BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

REMARKS

Claims 1-8 and 10-48 are pending in the present application, claim 9 being cancelled and new claims 13-48 added by the above amendment. The new claims are drawn to the same category of invention as the elected restriction group, so should be considered in the present application. Claims 10 and 11 were withdrawn by the Examiner after restriction, leaving claims 1-8 and 12-48 under consideration. Claims 1, 3, 4, 7, 8 and 12 are amended by the present submission. Support for these amendments and new claims can be found throughout the application as filed, including at page 2, lines 27-29; page 3, line 26 through page 4, line 19; page 5, lines 2-5 and 21-24; page 8, line 23 through page 9, line 1; page 9, lines 9-31; page 10, line 3-4; page 13, lines 8-18; page 13, line 29 through page 14, line 3; page 16, line 29 through page 19, line 2; page 22, line 6 through page 26, line 23; figure 14; the claims as filed; and the sequence listing, *inter alia*. No new matter has been added.

SEQ ID NO:1 is a cDNA encoding what applicants have designated "GTAR14-1 protein." The coding sequence corresponds to nucleotides 9-947 of SEQ ID NO:1 (see, e.g., page 16, lines 12-16, Figure 14, and the Sequence Listing). SEQ ID NO:4 is the amino acid sequence of the GTAR14-1 protein.

Objections to the Specification

The Examiner objected to the title of the application as not descriptive (Office Action, page 3). The title has been amended to "HUMAN GTAR G-PROTEIN COUPLED RECEPTOR" at the Examiner's suggestion.

Further, the Examiner objected to the specification as not fully in compliance with the sequence rules (Office Action, p. 3). Applicants respectfully direct the Examiner's attention to the amendments submitted on August 8, 2002 in response to the Notification of Missing Requirements, in which the specification was been amended to insert the sequence identifiers into the figure legends for figures 1-6, 10-19, and 24-31. Absent notification to the contrary, Applicants assume that the amendments contained therein were entered. The specification is further amended by the present submission to insert sequence identifiers for the primers shown in

Figures 4-6 and 16-19 into the appropriate figure legends. When a sequence is presented in a drawing, MPEP 2422.02 allows for the sequence identifier to be included in the Brief Description of the Drawings. Therefore, Applicants submit that the specification as amended does comport with the sequence rules, 37 CFR 1.821-1.825.

Applicants submit that the specification is now compliant and request withdrawal of the objections thereto.

Claim Objections

Claims 1 and 12 have been amended to remove all reference to non-elected subject matter, as requested by the Examiner. Applicants request withdrawal of the objection to these claims.

35 USC §§ 101 and 112, first paragraph

Claims 1-6 and 12 were rejected under 35 USC § 101 "because the claimed invention lacks a credible, specific and substantial utility." Office Action, p. 4. Applicants respectfully traverse.

The utility requirement is a relatively low threshold. From the MPEP, § 2107.1:

> ...as the Federal Circuit has stated, "[t]o violate [35 U.S.C.] 101 the claimed device **must be totally incapable of achieving a useful result**." Brooktree Corp. v. Advanced Micro Devices, Inc., 977 F.2d 1555, 1571, 24 USPQ2d 1401, 1412 (Fed. Cir. 1992) (emphasis added). See also E.I. du Pont De Nemours and Co. v. Berkley and Co., 620 F.2d 1247, 1260 n.17, 205 USPQ 1, 10 n.17 (8th Cir. 1980) ("A small degree of utility is sufficient . . . The **claimed invention must only be capable of performing some beneficial function** . . . An invention does not lack utility merely because the particular embodiment disclosed in the patent lacks perfection or performs crudely . . . A commercially successful product is not required . . . Nor is it essential that the invention accomplish all its intended functions . . . or operate under all conditions . . . partial success being sufficient to demonstrate patentable utility . . . **In short, the defense of non-utility cannot be sustained without proof of total incapacity**." If an invention is only partially successful in achieving a useful result, a rejection of the claimed invention as a whole based on a lack of utility is not appropriate. See In re Brana, 51 F.3d 1560, 34 USPQ2d 1436 (Fed. Cir. 1995); In re Gardner, 475 F.2d 1389, 177 USPQ 396 (CCPA), *reh'g denied*, 480 F.2d 879 (CCPA 1973); *In re Marzocchi*, 439 F.2d 220, 169 USPQ 367 (CCPA 1971).

Furthermore, the MPEP exhorts the Examiner to "focus on and be receptive to assertions made by the applicant that an invention is 'useful' for a particular reason." MPEP § 2107.01. The Guidelines for Examination of Applications for Compliance with the Utility Requirement (MPEP § 2107, referred to herein as "the Utility Guidelines") further state:

> Office personnel are reminded that they must treat as true a statement of fact made by an applicant in relation to an asserted utility, unless countervailing evidence can be provided that shows that one of ordinary skill in the art would have a legitimate basis to doubt the credibility of such a statement.

As further explained by the Utility Guidelines,

> If the applicant has asserted that the claimed invention is useful for any particular practical purpose (i.e., it has a "specific and substantial utility") and the assertion would be considered credible by a person of ordinary skill in the art, do not impose a rejection based on lack of utility. A claimed invention must have a specific and substantial utility. This requirement excludes "throw-away," "insubstantial," or "nonspecific" utilities, such as the use of a complex invention as landfill, as a way of satisfying the utility requirement of 35 U.S.C. 101.

The Applicants' assertions of utility hardly sink to the level of use as a landfill.

The present specification is rife with examples of the utility of the claimed invention. For example, the claimed GTAR14-1 nucleic acids and proteins exhibit tissue-specific expression patterns. mRNA for GTAR14-1 is observed only in the thymus during fetal development, and in testis, pancreas, and lymphatic/hematopoietic tissues including thymus and spleen in adults (see the specification at page 2, lines 32-35; page 6, lines 6-18; page 39, lines 22-25; and Fig. 7). All of these organs and tissues are known to be play a role in regulation of the lymphatic and endocrine systems. Thus, at the very least, one of skill in the art would appreciate that the claimed GTAR14-1 nucleic acids and polypeptides can be used as tissue markers (detected by, for example, immunohistochemical detection of antibodies generated against the claimed polypeptides or fragments thereof, or *in situ* hybridization using labeled fragments of the claimed polynucleotides) for those specific tissues, and can be used for developmental studies including cell fate experiments, for example, to evaluate the effects of a drug on the development of the immune system. This provides a well-established utility. To illustrate this point, Applicants

note that a similar approach can be seen in Forster et al., Blood 84(3):830-840 (1994) (abstract attached hereto as Exhibit A) which describes the use of BLR1 as a marker for memory T cells. The marker is detected using a BLR1-specific monoclonal antibody. Other examples include Cornish et al., B. J. Exp. Pathol. 70(5):495-504 (1989)) (abstract attached hereto as Exhibit B), which describes the use of antibodies to various B- and T-cell subsets to detect changes in the levels of those subsets associated with the induction of Heymann nephritis. Rocha et al., Blood 89(6):2189-202 (1997) (abstract attached hereto as Exhibit C), describe the use of monoclonal antibodies to distinguish donor from host cells, to differentiate between subsets of T lymphocytes, and stain certain macrophages, after systemic transfer of allogeneic antitumor immune T lymphocytes. Clearly, one of skill in the art would consider the similar information provided in the present specification a credible assertion of utility.

The tissue-specific expression pattern exhibited by GTAR14-1 provides a reasonable basis for a role of GTAR14-1 in immune function, including lymphoid development, e.g., the differentiation, proliferation or activation of lymphatic cells (see, e.g., page 6, lines 6-27). One of skill in the art would consider it reasonable and credible to suggest that GTAR14-1 is useful in screening for new targets and therapeutics for the modulation of immunity. The use of GTAR14-1 nucleic acids and polypeptides to screen for agents that modulate lymphatic/hematopoietic function or development is described in detail in the specification, see, e.g., pages 16-20. The specification at page 3, lines 21-25, describes the use of the claimed proteins in methods of screening for the natural ligand of GTAR14-1, i.e., a novel hematopoietic factor, or a novel hormone-like peptide factor. Both of these disclosures provide a substantial and specific utility.

The Examiner dismisses the use of the polypeptides in the generation of GTAR14-1 antibodies as "not specific or substantial... [I]f the specification discloses nothing specific and substantial about the polypeptide, both the polypeptide and its antibodies have no patentable utility." Office Action, page 6. As noted above, the GTAR14-1 polypeptides are useful at least as tissue markers; antibodies generated using the polypeptides as antigens are useful in detecting GTAR14-1 in tissues, and thus are useful in tissue identification and developmental analysis.

The claimed polypeptides, and antigenic fragments thereof, are useful to generate such antibodies, and additionally as positive controls for the antibodies.

The Examiner further makes the assertion that "[n]o well-established utility exists for newly isolated, complex biological molecules." Office Action, p 4. Applicants respectfully request that the Examiner explain what statutory or case law language supports such a sweepingly conclusory statement. If by "newly isolated" the Examiner means "uncharacterized," Applicants respectfully point out that this certainly does not describe the facts of the present case. If instead the Examiner means the term to apply to all novel genes and proteins, whether or not they are characterized, Applicants respectfully note that such a position is entirely inconsistent with U.S. practice. Genes and proteins are without question patentable under U.S. law. Clarification is requested so that Applicants can respond appropriately.

The Examiner is further reminded that

> The PTO has the initial burden of challenging a presumptively correct assertion of utility in the disclosure.... Only after the PTO provides evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility does the burden shift to the applicant to provide rebuttal evidence sufficient to convince such a person of the invention's asserted utility.

In re Brana, 51 F.3d 1560, 34 USPQ2d 1436 (Fed. Cir. 1995) (emphasis added). Applicants have provided sufficient characterization of the nucleotide and amino acid structure, significant structural homology to a relatively-well characterized group of proteins (the G-protein coupled odorant receptors), and expression pattern of these "newly isolated, complex biological molecules" to allow one of skill in the art to identify a number of possible uses, explicitly disclosed in the specification or not, that are certainly "well-established" enough to satisfy the requirements of 35 USC § 101. The Examiner has provided no evidence to contradict the Applicants' assertions of utility.

For the foregoing reasons, Applicants submit that the claimed invention possesses sufficient utility, and respectfully request withdrawal of the rejection under 35 USC § 101.

Claims 1-9 and 12 were rejected under 35 USC § 112, first paragraph, "since the claimed invention is not supported by either a specific and substantial asserted utility or a well established utility for the reasons set forth above, one skilled in the art clearly would not know how to use the claimed invention." Office Action, pages 7-8. Claim 9 has been cancelled. Applicants respectfully traverse the rejection as it may apply to the remaining claims.

For the reasons given above, Applicants submit that the claimed invention does possess utility. In addition, the specification provides substantial guidance as to how to use the invention as presently claimed. See, e.g., Example 1, which describes the isolation and analysis of GTAR14-1. The specification teaches the sequence (i.e., the primary structure) of the GTAR14-1 polypeptide, and of a nucleic acid encoding that polypeptide. Methods of making recombinant polypeptides, and methods of using the polypeptides for the generation of antibodies, are described in detail. The references cited by the Examiner deal with the difficulties of predicting function based solely on homology to an annotated sequence in a database. The predictions of function included in the present application are not based solely on homology to annotated sequences; first, the sequences to which GTAR14-1 has homology are proteins with proven function, members of a described family, not simply sequences randomly grouped by automated correlation software. GTAR14-1 has significant homology to members of the olfactory receptor gene family (see, e.g., Fig, 1), has seven transmembrane regions characteristic of G-protein coupled receptors (see Figs. 10 and 11), and has amino acids known to be essential for function of G-protein coupled receptors (see page 5, line 10 to page 6, line 1). The data regarding the function of these homologs comes not from database annotation, but from empirical observations of the function of the olfactory receptors. Furthermore, the predictions of function are supported by details of the tissue-restricted expression patterns of GTAR14-1. One of skill in the art, given this information and the other information included in the application as filed, would be able to use the claimed polypeptides and polynucleotides, e.g., as probes to detect tissue-specific markers, and in screens for novel hematopoietic factors, or novel hormone-like peptide factors, e.g., modulators of immunity.

The Examiner additionally stated that "the specification does not reasonably provide enablement for all *fragments and variants* of SEQ ID NO:4" (Office Action, page 10; italics in original). Claim 3, which referred to functional fragments of the protein of claim 1, has been amended to refer to antigenic fragments of the polypeptide of claim 14: i.e., SEQ ID NO:4. Claim 1 has been amended to claim an isolated G protein-coupled receptor protein selected from the group consisting of (a) a polypeptide comprising the amino acid sequence of SEQ ID NO:4; and (b) a polypeptide comprising the amino acid sequence of SEQ ID NO:4, in which one to thirty amino acids are deleted, added, inserted, or substituted by another amino acid residue. This substantially limits the fragments and variants, both structurally and by function (since each protein covered by claim 1 must have the function of a G-protein coupled receptor). Applicants submit that the requirements of 35 U.S.C. § 101 and § 112, second paragraph, are amply satisfied.

35 USC § 112, First Paragraph – Written Description

Claims 1-6 were rejected as "containing subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention." Office Action, pages 11-12. In particular, the Examiner stated that:

> The specification as originally filed does not provide adequate written description for an isolated protein wherein one to thirty amino acids of the receptor protein are modified or deleted, while still maintaining the function of the receptor.

Office Action, page 12. The Applicants respectfully traverse and direct the Examiner's attention to the specification as originally filed at page 8, line 11 through page 9, line 1, especially page 8, line 25, as well as claim 1 as originally filed. The rejection seems to be based on the fact that the applicants did not disclose the results of functional assays on the GTAR14-1 protein. Applicants note that the sole function recited in claim 1 is "G-protein coupled receptor protein." This general class of receptor proteins has been extensively studied in the art. A great deal is known about the structure and function of various structural domains of receptors in this class. Thus, one of ordinary skill in the art would know that certain residues of SEQ ID NO:4 (e.g.,

hydrophobic residues in one of the seven transmembrane domains characteristic of the class) can be substituted with similar residues, with little likelihood of losing the "G-protein coupled receptor" function of the protein (see, e.g., the pre-October, 1998 studies reviewed in Gershengorn and Osman, Endocrinology 142(1):2-10 (2001) (attached hereto as Exhibit D); also see Ji et al., J. Biol. Chem. 273(28):17299-17302 (1998) (attached hereto as Exhibit E)). Likewise, it is known that intracellular domains of different G-protein coupled receptors can be swapped, with the resulting hybrid receptors retaining their general G-protein coupled receptor characteristics. (See Postina et al., Adv. Exp. Med. Biol. 449:371-85 (1998) (abstract attached hereto as Exhibit F)).

It is generally known that proteins are "highly tolerant" of amino acid changes (see, e.g., Bowie et al. (1990) Science 247:1306-1310 (copy enclosed as Exhibit G). At page 1306, lines 12-13, Bowie teaches that "proteins are surprisingly tolerant of amino acid substitutions." Bowie et al. cites as evidence a study carried out on the lac repressor. Of approximately 1500 single amino acid substitutions at 142 positions in this protein, about one-half of the substitutions were found to be "phenotypically silent," i.e., had no noticeable effect on the activity of the protein (Bowie, at page 1306, col. 2, lines 14-17). Presumably the other half of the substitutions exhibited effects ranging from slight to complete abolishment of repressor activity. Thus, one can expect, based on Bowie et al.'s teachings, to find over half (and possibly well over half) of random substitutions in any given protein to result in proteins with full or nearly full activity. These are far better odds than those at issue in In re Wands, 858 F.2d 731 (Fed. Cir. 1988), in which the court said that screening many hybridomas to find the few that fell within the claims was not undue experimentation. Based on Bowie et al.'s teachings, one would predict that even random deletions, additions, insertions, or substitutions of residues in SEQ ID NO:4 will predictably result in a majority of the variants having full or partial receptor activity. Given what is known about G protein coupled receptors in general, one of ordinary skill in the art could readily make intelligent decisions about what positions to alter and which to leave untouched in order to preserve the claimed function. The Examiner has presented no evidence to the contrary. Applicants submit that the claims satisfy the requirements of 35 USC § 112, first paragraph.

Claims 8 and 9 were also rejected by the Examiner for lack of written description. Claim 9 has been cancelled, obviating the rejection as to that claim. Claim 8 recites a method of screening for a compound that binds to a G protein-coupled receptor protein, the method comprising (a) exposing a test sample to the protein of claim 1, and (b) selecting a compound that binds to the protein. New claims 41-43 further specify the nature of the protein used in the method of claim 8 (in claim 41, the protein comprises SEQ ID NO:4, in claim 42, the protein consists of SEQ ID NO:4, and in claim 43, the protein is a fusion protein comprising SEQ ID NO:4 and another peptide or polypeptide). Ample support is found in the specification for the use of the claimed proteins in simple binding assays; see, e.g., page 16, line 29 to page 20, line 4, which describes methods that include Western blotting, yeast two-hybrid assays, and affinity chromatography. None of these methods, nor claim 8 (or 41-43), requires specific knowledge of the function of the protein.

The Examiner stated:

> The specification as originally filed does not provide adequate written description for a method of using a receptor protein with an unknown function to search for specific ligands. This limitation is not expressly asserted, nor does it flow naturally from the specification as originally presented.

Office Action, page 13. Applicants respectfully traverse, as the specification does expressly disclose methods of assaying for ligands without knowing the specific function of the receptor protein. See, e.g., page 16, line 29 to page 20, line 4, describing methods including immunoprecipitation, yeast two-hybrid, and Western blotting. None of those methods require any foreknowledge of the specific function of the protein. As mentioned on page 17, lines 5-14, the bound compound may be the receptor's ligand.

Finally, the Examiner further stated that

> The specification as originally filed does not provide adequate written description of the method now claimed. The specification teaches the receptor polypeptide of SEQ ID NO:4. However, functional assays of the receptor were not performed. It is not known what class of ligands to use, what tests to perform, nor what tissues to use, for example.

Office Action, page 13. Applicants point out that the written description requirement has never required that the claimed method have been "performed" in order to be patentable. Nor is it necessary to disclose the "class of ligands" to use, as that "class" is not limited in the claim. Any class of compound can be screened. (Though in that regard, Applicants point out that the specification does describe several possible categories of substances to screen: "peptides, purified or crude preparation of proteins, non-peptide compounds, synthetic compounds, products of microorganism fermentation, cell extracts, animal tissue extracts, marine organism extracts, and plant extracts" (see page 19, lines 17-21). As described above, ample description of "what tests to perform" can be found in the specification. For example, as noted above, the disclosure at page 16, line 29, to page 20, line 4, describes methods that include Western blotting, yeast two-hybrid assays, and affinity chromatography. Tissues that can be used are also described, as the tissues in which GTAR14-1 is expressed are amply described in the specification (see, e.g., page 2, lines 32-35; page 6, lines 6-18; page 39, lines 22-25; and Fig. 7), as are any number of heterologous expression systems (see, e.g., page 10, line 22 to page 14, line 34). For the foregoing reasons, Applicants submit that claims 8 and 41-43 satisfy the written description requirement of 35 USC § 112, first paragraph.

35 USC § 112, Second Paragraph - Indefiniteness

Finally, the Examiner rejected claim 12 as being indefinite for the use of the term "under highly stringent conditions." The Examiner indicated that this rejection could be overcome by supplying specific conditions, supported by the specification, which the Applicants consider "stringent." Office Action, page 14. Claim 12 has been amended to recite "hybridizes at 65°C in 2x SSC and 0.1% SDS." Applicants submit that this is sufficient to overcome the rejection under 35 USC § 112, second paragraph.

For the foregoing reasons, Applicants submit that the claims as amended satisfy the requirements of 35 USC §§ 101 and 112, and respectfully request withdrawal of the pending rejections and allowance of the claims as amended. If the Examiner feels that a telephone

conference would be useful to expedite prosecution, she is hereby invited to telephone the undersigned at 617-956-5985.

Enclosed is a check for the Petition for Extension of Time fee and a check for the Excess Claim fee. Please apply any other charges or credits to deposit account 06-1050, referencing attorney docket no. 14875-075001.

Respectfully submitted,

Date: 2·20·2004

Janice L. Kugler
Reg. No. 50,429

Fish & Richardson P.C.
225 Franklin Street
Boston, MA 02110-2804
Telephone: (617) 542-5070
Facsimile: (617) 542-8906

20742779.doc

1: Blood. 1994 Aug 1; 84(3): 830-40.

Expression of the G-protein--coupled receptor BLR1 defines mature, recirculating B cells and a subset of T-helper memory cells.

Forster R, Emrich T, Kremmer E, Lipp M.

Max-Delbruck-Center for Molecular Medicine, Berlin-Buch, Germany.

The G-protein-coupled receptor BLR1 related to receptors for chemokines and neuropeptides has been identified as the first lymphocyte-specific member of the gene family characterized by seven transmembrane-spanning regions. Using a high-affinity anti-BLR1 monoclonal antibody (MoAb) and three-color flow cytometry it is shown that BLR1 expression on peripheral blood cells is limited to B cells and to a subset of CD4+ (14%) and CD8+ (2%) lymphocytes. T cells expressing BLR1 were positive for CD45R0, were negative for interleukin-2 receptors, show high levels of CD44, and show low levels of L-selectin. The majority of CD4+ cells originating from secondary lymphatic tissue, but none of cord blood-derived T cells, express BLR1. These observations suggest that BLR1 is a marker for memory T cells. Furthermore, BLR1 expression was detected on all CD19+ peripheral or tonsillar B lymphocytes, but only on a fraction of cord blood cells and bone marrow cells expressing CD19, sIgM, or sIgD. Interestingly, activation of both mature B and T cells by CD40 MoAb and CD3 MoAb, respectively, led to complete downregulation of BLR1. These data suggest that the G-protein-coupled receptor BLR1 is involved in functional control of mature recirculating B cells and T-helper memory cells participating in cell migration and cell activation.

PMID: 7913842 [PubMed - indexed for MEDLINE]

1: Br J Exp Pathol. 1989 Oct; 70(5): 495-504.

Immunoregulation in Heymann nephritis. I. Cell marker studies.

Cornish J, Barabas AZ, Lannigan R, Rozing J.

Department of Pathology, University of Calgary, Alberta, Canada.

Immunoregulation was examined in rats with Heymann nephritis (HN), an established model of membranous glomerulonephropathy (MGN). There is little known of the cellular immune events for the induction and maintenance of the autoimmune response in HN. The cell marker studies utilized fluorescein (FITC)-labelled monoclonal antibodies directed to B cells (Mark-I), and T cell subsets: pan T (ER-I), helper/inducer T (ER-2) and suppressor/cytotoxic T (ER-3). Lymphoid subsets were compared in spleen, lymph nodes, peripheral blood and bone marrow, of normal and diseased rats. Animals were investigated during the induction and chronic phases of disease. The induction of HN was associated with an early, significant, but transient increase of the non-specific myeloid component of the defence system. Subsequently, a significant increase was seen in the number of cells of the B lymphocyte lineage in HN animals, which coincided well with the overall increased humoral immune responsiveness. No alterations in the T lymphocyte subsets were noted during the development of this experimental autoimmune disease.

PMID: 2684253 [PubMed - indexed for MEDLINE]

☐ 1: Blood. 1997 Mar 15; 89(6): 2189-202.

Differences between graft-versus-leukemia and graft-versus-host reactivity. I. Interaction of donor immune T cells with tumor and/or host cells.

Rocha M, Umansky V, Lee KH, Hacker HJ, Benner A, Schirrmacher V.

Deutsches Krebsforschungszentrum Heidelberg, Abteilung Zellulare Immunologie, FSP Tumorimmunologie, Germany.

Graft-versus-leukemia (GVL) and Graft-versus-host (GVH) reactions were compared after systemic transfer of allogeneic antitumor immune T lymphocytes from B10.D2 (H-2d; Mls (b)) into DBA/2 (H-2d; Mis(a)) mice. Before immune cell transfer, recipient DBA/2 mice were sublethally irradiated with 5 Gy to prevent host-versus-graft reactivity. Recipients were either bearing syngeneic metastatic ESb lymphomas (GVL system) or were normal, non-tumor-bearing mice (GVH system). We previously reported that this adoptive immunotherapy protocol (ADI) had pronounced GVL activity and led to immune rejection of even advanced metastasized cancer. In this study, monoclonal antibodies were used for immunohistochemical analysis of native frozen tissue sections from either spleen or liver to distinguish donor from host cells, to differentiate between CD4 and CD8 T lymphocytes, and to stain sialoadhesin-positive macrophages at different time points after cell transfer. The kinetics of donor cell infiltration in spleen and liver differed in that the lymphoid organ was infiltrated earlier (days 1 to 5 after transfer) than the nonlymphoid organ (days 5 to 20). After reaching a peak, donor cell infiltration decreased gradually and was not detectable in the spleen after day 20 and in the liver after day 30. The organ-infiltrating donor immune cells were mostly T lymphocytes and stained positive for CD4 or CD8 T-cell markers. A remarkable GVL-associated observation was made with regard to a subset of macrophages bearing the adhesion molecule sialoadhesin (SER+ macrophages). In the livers of tumor-bearing mice, their numbers increased between days 1 and 12 after ADI by a factor greater than 30. Double-staining for donor cell marker and SER showed that the sialoadhesin-expressing macrophages were of host origin. The SER+ host macrophages from GVL livers were isolated by enzyme perfusion and rosetting 12 days after ADI, when they reached peak values of about 60 cells per liver lobule, and were tested, without further antigen addition, for their capacity to stimulate an antitumor CD8 T-cell response. The results of this immunologic analysis suggest that these cells in the liver function as scavengers of the destroyed metastases and as antigen-processing and -presenting cells for antitumor immune T cells.

PMID: 9058744 [PubMed - indexed for MEDLINE]

EXHIBIT D

0013-7227/01/$03.00/0
Endocrinology

Vol. 142, No. 1
Printed in U.S.A.

Minireview: Insights into G Protein-Coupled Receptor Function Using Molecular Models*

MARVIN C. GERSHENGORN AND ROMAN OSMAN

Division of Molecular Medicine (M.C.G.), Department of Medicine, Weill Medical College and Graduate School of Medical Sciences of Cornell University, New York, New York 10021; and Department of Physiology and Biophysics (R.O.), Mount Sinai School of Medicine, New York, New York 10029

ABSTRACT

G protein-coupled receptors (GPCRs) represent the largest family of signal-transducing molecules known. They convey signals for light and many extracellular regulatory molecules. GPCRs have been found to be dysfunctional/dysregulated in a growing number of human diseases and have been estimated to be the targets of more than 30% of the drugs used in clinical medicine today. Thus, understanding how GPCRs function at the molecular level is an important goal of biological research. In order to understand function at this level, it is necessary to delineate the 3D structure of these receptors. Recently, the 3D structure of rhodopsin has been resolved, but in the absence of experimentally determined 3D structures of other GPCRs, a powerful approach is to construct a theoretical model for the receptor and refine it based on experimental results. Computer-generated models for many GPCRs have been constructed. In this article, we will review these studies. We will place the greatest emphasis on an iterative, bi-directional approach in which models are used to generate hypotheses that are tested by experimentation and the experimental findings are, in turn, used to refine the model. The success of this approach is due to the synergistic interaction between theory and experiment. (*Endocrinology* **142:** 2–10, 2001)

G PROTEIN-COUPLED receptors (GPCRs) represent the largest family of signal-transducing molecules known. For example, GPCRs comprise more than 4% of the genes in *Caenorhabditis elegans*. GPCRs convey signals for light and many extracellular regulatory molecules, such as, hormones, growth factors, and neurotransmitters, that regulate every cell in the body. Dysregulation of GPCRs has been found in a growing number of human diseases, and GPCRs have been estimated to be the targets of more than 30% of the drugs used in clinical medicine today. Thus, understanding how GPCRs function at the molecular level is an important goal of biological research.

GPCRs may be grouped into three subfamilies that exhibit little sequence homology but appear to share the same overall topology. In this review, we will focus on receptors in the largest subfamily, which comprises more than 90% of GPCRs, that includes rhodopsin (Rh) and receptors for small neurotransmitters, such as opioids and catecholamines, peptides such as bradykinin, GnRH, and TRH, and glycoprotein hormones. The two-dimensional topology of a typical member of this subfamily, a receptor for TRH, is illustrated in Fig. 1[1]. GPCRs contain an extracellular amino terminus, three extracellular loops (ECLs), seven transmembrane-spanning helices (TMHs), three intracellular loops (ICLs) and an intracellular carboxyl terminus. The specific domains within which different regulatory molecules bind to GPCRs, and the changes in a GPCR that constitute conversion of the receptor from an inactive to an active state are under intense investigation. In general, evidence has been presented that small ligands bind primarily in the core of the TMHs, intermediate size peptides to the ECLs and TMHs, and larger peptides and proteins to the amino terminus, ECLs and, most likely, the TMHs of their specific receptors. Upon binding an agonist, it is thought that changes are induced in the conformation of the receptor, in particular, changes involving the relative positions of the seven TMHs. The movements of the TMHs are thought to cause changes in the ICLs leading to increased coupling to a heterotrimeric G protein(s) and G protein activation. These hypotheses, however, have only begun to be tested by direct experimentation.

Recently (1), the three-dimensional (3D) structure of rhodopsin has been resolved. In the absence of experimentally determined 3D structures for other GPCRs, one approach is to construct a model for the GPCR and to refine the model based on experimental results. Computer-generated models for many GPCRs have been constructed. A number of these models are available at http://swift.embl-heidelberg.de/7tm/models/lin/index.html. The reliability of the predicted model is difficult to assess independently. However, its usefulness is in providing a molecular basis for specific structural hypotheses with concomitant functional correlates. Moreover, even when 3D structures are resolved experimentally, molecular models will be useful to generate hypotheses regarding the molecular details of GPCR function. Models of the TMH domains are constructed on the basis of a topological homology, expressed as a general backbone template (see below). However, the specific sequence of the receptor

Received August 28, 2000.

Address all correspondence and requests for reprints to: Dr. Marvin C. Gershengorn, Weill Medical College of Cornell University, 1300 York Avenue, Room A328, New York, New York 10021-4896. E-mail: mcgersh@med.cornell.edu.

* The work from our laboratories was supported by USPHS Grant DK-43036.

[1] In this review, we elected to use the TRH receptor as the only GPCR for which the primary sequence is shown. Therefore, only for TRH receptor will the positions of specific residues be identified. For all other GPCRs, specific residues will be described as being part of a conserved sequence/motif or as part of a particular receptor domain.

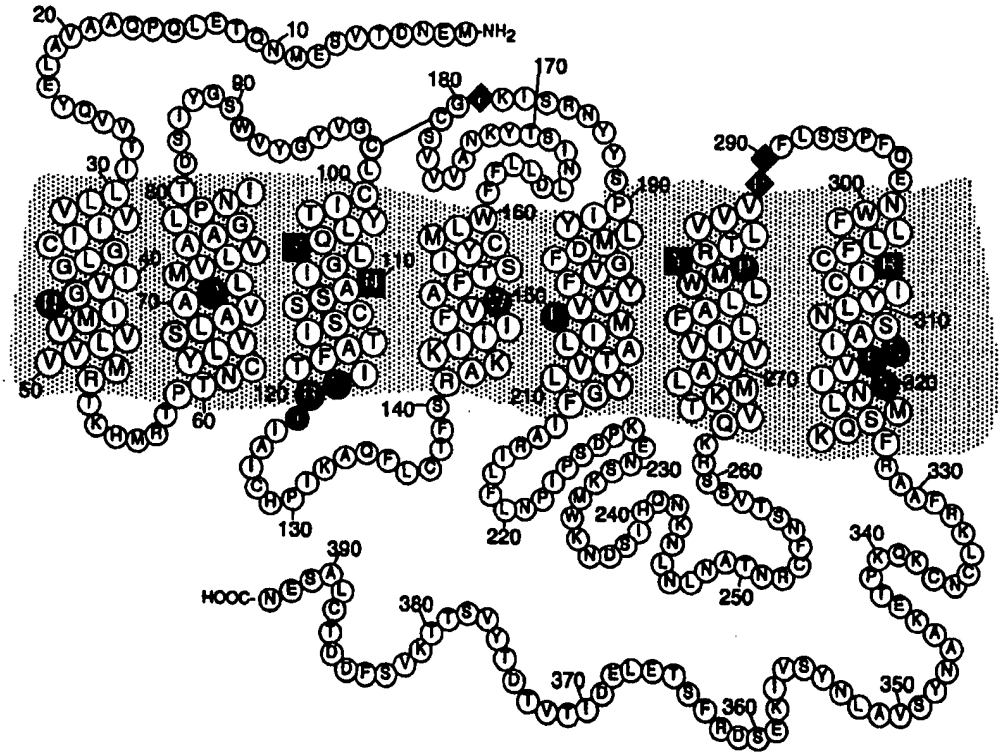

FIG. 1. A two-dimensional topology of the TRH receptor sequence. Two-dimensional topology of the TRH receptor with the extracellular space at the top and the intracellular space at the bottom. The residues highlighted in circles are sites of high conservation that were used to determine the helical boundaries and align the helices on the transmembrane template. The residues highlighted in squares constitute the transmembrane binding pocket and those highlighted in diamonds the residues that participate in anchoring TRH to the receptor surface and guiding it to the transmembrane pocket.

determines local structure and interactions between proximal residues. Therefore, the initial structure of a predicted model must be refined computationally by minimization of the initial structure to relieve the offending close contacts. Subsequently, dynamic simulations that efficiently sample configurational space are used to refine the structure. The construction of loops in receptors encounters the difficulty of no available templates. Here, a simulated annealing approach is necessary to ensure proper folding of the loops and constant temperature simulations can be used to refine the initial fold.

Some investigators used bacteriorhodopsin (Br), for which a crystal structure at 1.55 Å resolution is now available (2), as a template for constructing the TMHs of GPCRs (3). Br, however, is not a GPCR even though it is a seven transmembrane-spanning protein. Other investigators used a GPCR template constructed by Baldwin (4). This TMH template was based on a comprehensive analysis of hundreds of GPCR sequences and was shown to be consistent with many mutational analyses. This template agrees well with the recently reported crystal structure of Rh that was resolved to 2.8 Å (1). It is now clear that position and tilt of TMHs in Rh, and by inference in all GPCRs, are different from those in Br. Nevertheless, models based on both Rh and Br have proven useful in generating hypotheses that have been tested experimentally.

In the majority of reports that describe computer simulations of GPCRs, the purpose of model construction was to describe the binding pocket for the ligand within the receptor. Models of ligand-receptor complexes were constructed incorporating experimental evidence of the specific receptor amino acid residues with which the ligands were found to interact. These models are being used, for example, in efforts to discover GPCR agonist or antagonist drugs. Another important use of GPCR models is to generate hypotheses regarding the binding and signaling functions of these receptors for experimental testing. That is, to direct experimentation to define the structure-function relationships of GPCRs. We think the most effective use of models, however, is as part of an iterative, bidirectional approach in which models are used to generate hypotheses that are tested by experimentation and the experimental findings are, in turn, used to revise and refine the model. The success of this approach is due to the synergistic interaction between theory and experiment. We will emphasize this last approach in this review.

Initial Models of Ligand-GPCR Complexes

Small ligands

The prototypical complex for a small ligand and its receptor is the 11-*cis*-retinal/Rh complex. In contrast to all other ligand/receptor complexes to be described, 11-*cis*-retinal is covalently attached to Rh. The linkage is via a protonated Schiff base to a Lys in TMH7. As noted above,

11-*cis*-retinal/Rh is the only ligand/GPCR complex for which a structure has been resolved experimentally (1). The Lys in TMH7 that is attached to retinal lies deep within the bundle of TMHs and the entire ligand is within the TMH bundle with the six-membered β-ionone ring closer to the extracellular aspect of Rh. The Schiff base is surrounded by residues from TMH1, 2, and 7 and the β-ionone ring interacts with residues from TMH3 and 6. Models of ligand/receptor complexes for a number of small ligands and their cognate GPCRs have been constructed (Table 1). The TMH bundle of the receptor was based on either the Rh or Br template. In every case, the ligand was docked within the receptor based on experimental results that identified receptor amino acid residues which directly interact with the ligand. That is, in no case was a stable model of a ligand/receptor complex constructed using a computer simulation independent of experimental evidence delineating the binding pocket. (This is true for ligand/GPCR complexes of peptide and protein ligands also.) A typical model of a GPCR for a small ligand, TRH (pyroGlu-His-ProNH_2), is illustrated in Fig. 2 and a close-up of the TRH/receptor complex is shown in Fig. 3. The details of our construction have been described elsewhere (5). TRH is predicted to interact with residues in TMH3, 6 and 7 within the extracellular part of the TMH bundle (see below).

Peptide ligands

As noted above, peptide ligands bind to GPCRs by interacting with the extracellular domains and TMHs. It is noteworthy that a disulfide bond links ECL1 to ECL2 (see Fig. 1). A typical example of a model of a peptide ligand/receptor complex is that of cholecystokinin-8, an octapeptide that is the most biologically active form of this peptide, with its receptor (6, 7). The interactions of cholecystokinin-8 with its receptor are predicted to involve residues in the amino terminus, ECL2, ECL3, and the extracellular aspect of TMH6. The nonapeptide arginine-vasopressin is predicted to bind to its cognate receptor by interacting with residues in ECL1 and TMH2, 3, 4, 6, and 7 (8, 9) and the decapeptide GnRH is predicted to bind to residues in all TMHs and ECL2 and 3 in

TABLE 1. Models of GPCRs and ligand/GPCR complexes[a]

Receptors for small ligands	References
Adenosine	(51,52)
Adrenergic	(40,46,53)
Cannabinoid	(54)
Dopamine	(55)
Formyl peptide	(56)
Histamine	(22,23)
Muscarinic	(57)
Opioid	(24,58)
Platelet-activating factor	(59)
Purinergic	(28,60)
Serotonin	(25,49)
TRH	(18–21,26,27,37,47,50)
Receptors for peptides	
Angiotensin	(61,62)
Bradykinin	(12,63)
Cholecystokinin	(6,7,30)
Endothelin	(64)
GnRH	(10,11,35,36)
Gn secretagog	(31)
Melanocortin	(65)
Melanin-concentrating hormone	(32)
Neuropeptide Y	(66)
Neurotensin	(67)
Oxytocin	(43)
Tachykinin	(68)
Vasopressin	(8,9,33,34)
Receptors for large polypeptides/proteins	
Chemokines	(13,14)
FSH	(69)
LH/CG	(15–17,48)

[a] Due to space limitations, we were only able to provide a limited number of references for the models listed and could not include all models for GPCRs that have been constructed.

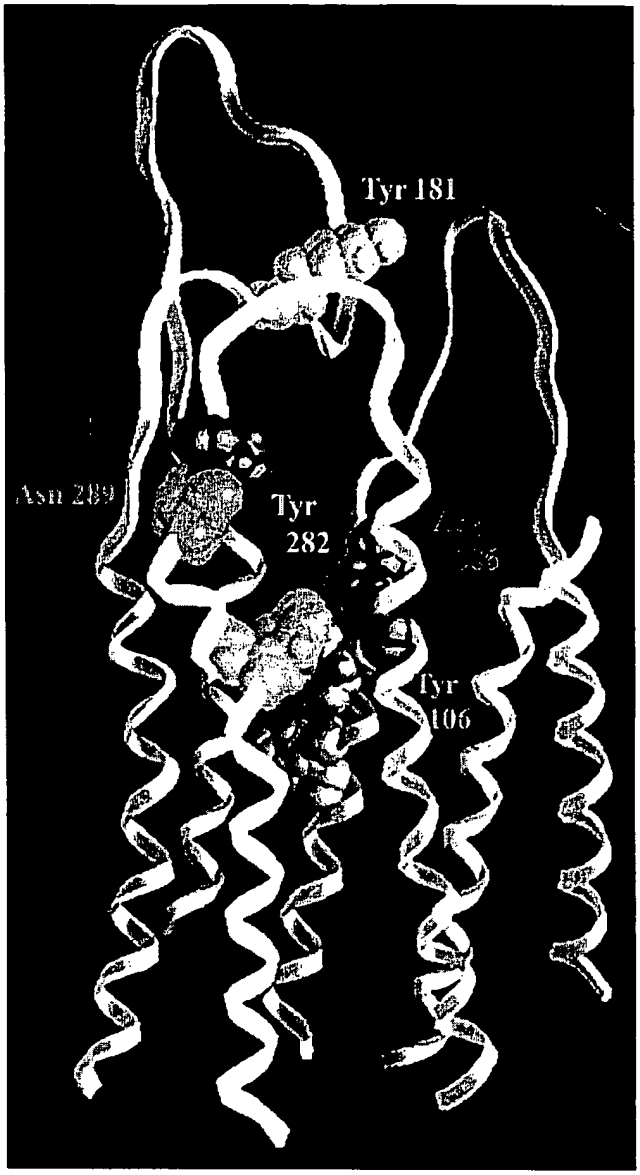

FIG. 2. Model of TRH receptor. A side view of the three-dimensional model of the TRH receptor with the extracellular space at the top and the intracellular space at the bottom. The model includes the extracellular loops and the transmembrane helices. The residues shown are those involved in TRH binding.

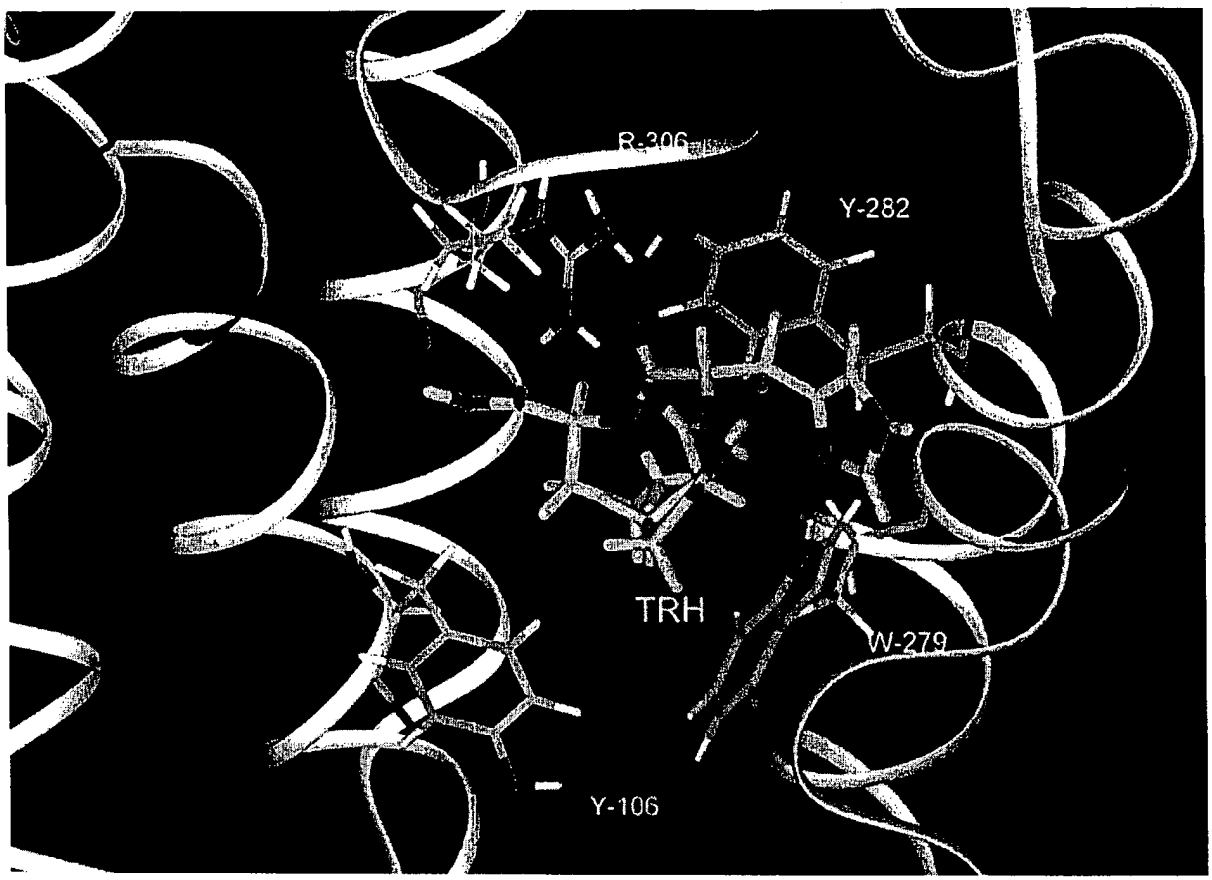

FIG. 3. A close-up of the transmembrane binding pocket of the TRH receptor showing TRH (*yellow*) in its bound conformation. TRH interacts with Tyr-106 in TMH3, Tyr-282 and Trp-279 in TMH6 and Arg-306 in TMH7.

its cognate receptor (10, 11). A model for the binding of the nonapeptide bradykinin predicts that bradykinin binds to residues in TMH5, 6, and 7 near the TMH/ECL boundary, whereas a peptide antagonist binds to residues in TMH5, 6, and 7 that are deeper within the TMH bundle (12). Different sites for binding peptide agonists, peptide antagonists and nonpeptide antagonists have been reported in a number of GPCRs.

Large polypeptide/protein ligands

As noted above, larger peptides and proteins bind to the amino terminus, ECLs and, most likely, the TMHs of their specific receptors. Chemokines are 70 to 100 amino acid polypeptides that bind primarily to the amino termini and ECL1 and 2 of their cognate receptors (13, 14). Models of receptors for glycoprotein hormones do not include the amino termini (15, 16), but it has been predicted that these ligands bind primarily to the amino termini and ECLs of their receptors (17).

Use of Models to Generate Hypotheses of GPCR Structure and Function

Small ligand binding pockets

We have constructed a model of the TRH receptor on a generic template for GPCRs (5). The generic template was developed based on an analysis of GPCR sequences and the constructed model was optimized using energy minimization and molecular dynamic simulations. We tested the modeling approach by constructing a model of Rh that we showed agreed well with a density projection map of the protein and with structural inferences derived from biochemical and mutational experiments. Highly conserved residues in all helices (Asn-43 in TM-1, Asp-71 in TM-2, Ser-112 in TM-3, Trp-150 in TM-4, Pro-203 in TM-5, Trp-279 in TM-6 and Asn-316 in TM-7) (Fig. 1) were used to guide the initial construction. In our initial model of the TRH/TRH receptor complex, we docked TRH into the TMH bundle of TRH receptor and, using a novel simulation approach, optimized the complex by extensive sampling of the configurational and conformational degrees of freedom (18). This model predicted that, in addition to the experimentally determined interaction of pyroGlu with Tyr-106 and Asn-110 in TMH3, Tyr-282 in TMH6, and Arg-306 in TMH7 could interact with the His and ProNH_2 of TRH, respectively. These predictions were tested using synthetic analogs of TRH and mutational analysis of TRH receptor, and the experimental observations were shown to be consistent with the predictions (19–21). The final model shows that the pyroGlu moiety of TRH forms H-bonds with Tyr-106 and Asn-110 in TMH3, His interacts

hydrophobically with Tyr-282 in TMH6, and the $ProNH_2$ forms H-bonds with Arg-306 in TMH7. Thus, these studies allowed us to describe the molecular interactions of TRH with the binding pocket within the TMH bundle of TRH receptor.

Ter Laak, Timmerman, Leurs, and colleagues (22, 23) constructed a model of an agonist/histamine H_1 receptor complex based on a Br TMH template using experimental evidence for the interactions between histamine agonists and specific receptor residues. The model predicted an Asp in TMH3, a Trp in TMH4, a Lys in TMH5 and two Phe residues in TMH6 with which ligands would interact. Experiments involving functional studies with mutated receptors and different antagonists provided evidence that the predictions were correct. These studies allowed delineation of agonist and antagonist binding pockets in histamine H1 receptor.

Befort and colleagues (24) constructed a model of the TMH bundle of the δ-opioid receptor based on a Br template and sequence alignment with other GPCRs. The model predicted that a series of aromatic residues in TMHs project into the bundle. These residues were mutated and binding of opioid analogs was measured. Evidence was obtained that a Tyr in TMH3, a Trp in TMH4, a Phe in TMH5, a Trp in TMH6, and a Tyr in TMH7 provide interactions with the ligands. These findings further demonstrated the importance of aromatic residues in binding small ligands.

Almaula and co-workers (25) constructed a model of the serotonin type 2A receptor that predicted that the ligand would interact with a Ser one helical turn below the Asp in TMH3 that binds the cationic primary amino group of serotonin. This prediction was shown to be consistent with data from experiments using Ser to Ala and to Cys mutations and serotonin analogs.

Sequential binding of small ligands to GPCRs

Based on our model of the unoccupied TRH receptor (26, 27) (Fig. 4), we proposed that TRH binds to TRH receptor in at least three discrete steps. TRH interacts with: 1) Tyr-181 on the surface of the ECLs; 2) with Asn-289 and Ser-290 at the ECL/TMH boundary; and 3) with TMH residues (see above). Specifically, it was predicted that the pyroGlu of TRH interacts with Tyr-181 then with Asn-289/Ser-290 and then with Tyr-106/Asn-110. These predictions were tested experimentally. Using a mutation of Tyr-181 to Phe, we showed that TRH interacted with Tyr-181 via the pyroGlu moiety because a mutant receptor in which Phe was substituted for Tyr-181 (Y181F) was not able to discriminate between TRH analogs changed in the pyroGlu position but was able to distinguish between TRH and analogs substituted in the His or $ProNH_2$ positions. Similar findings were obtained with Y106F and N110A mutant TRH receptors, and with N289A and S290A mutant receptors. Most importantly, N289A exhibited a decreased binding affinity for TRH that was accounted for by a decreased rate of association rather than N110A that exhibited a similar decrease in affinity but in which the decreased affinity was caused by an *increased* rate of dissociation (27). Thus, we concluded that TRH is attracted to TRH receptor via interactions with Tyr-181, is conducted down a funnel formed by the ECLs through which it enters the TMH binding pocket in which it exerts its effect to change the conformation of TRH receptor that comprises the conversion of the receptor from an inactive to active state (see below).

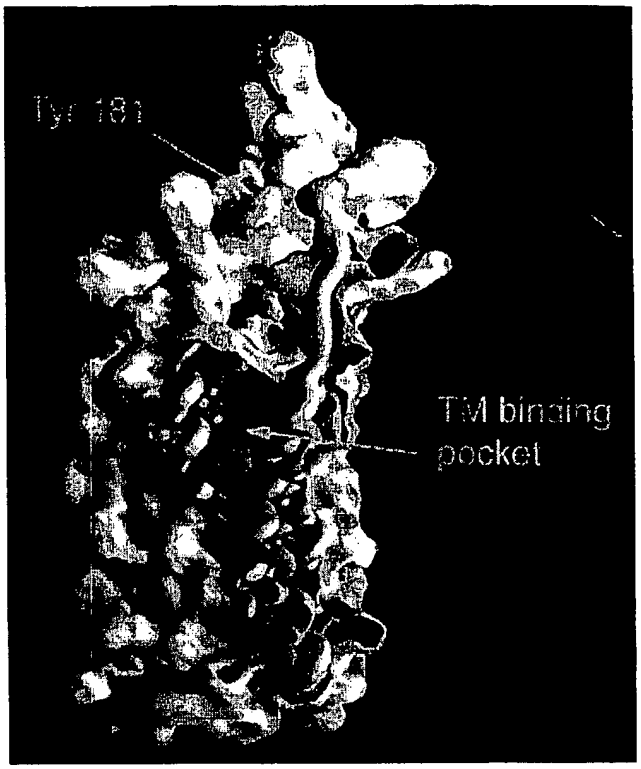

FIG. 4. A depiction of the funnel formed by the extracellular loops that guides TRH into the transmembrane binding pocket. The surface has been partially removed to show the funnel, with Tyr-181 at its entrance, as well as the transmembrane (TM) binding pocket. This figure is based on recent simulation conducted by Dr. Rosenhouse-Dantsker.

Using molecular modeling and experiments with mutant receptors and purinergic analogs, Moro and colleagues (28) drew similar conclusions regarding sequential binding of purinergic ligands to the $P2Y_1$ receptor.

Thus, it appears that small ligands enter their binding pockets within the TMH bundle after attraction and sequential interactions with ECL residues.

Peptide ligand binding pockets

Several groups of investigators have used initial models of ligand/GPCR complexes to make predictions regarding other interactions between receptor residues and ligands to define the binding pockets of receptors for peptides more completely. The binding pockets for nonpeptide ligands in these same receptors were analyzed also. In a series of reports (6, 7, 29), Fourmy and colleagues described experiments in which they used a series of cholecystokinin (CCK) analogs and cholecystokinin-A receptor (CCK-AR) mutants coupled with computer simulations to develop a model of CCK binding and CCK-AR activation. An initial model of CCK-AR was constructed based on the TMH template of Br, and Rh and β2-adrenergic receptor for sequence alignment. Based on experimental findings that two residues in the amino termi-

nus of CCK-AR bind the amino terminus of CCK-9 and docking the remainder of the CCK-9 peptide into the TMH bundle of CCK-AR, a model of the CCK/CCK-AR complex was constructed. (The 3D structure of CCK-9 was based on biophysical measurements.) The model was optimized by energy minimization and molecular dynamics simulation protocols. This model predicted that a Met residue in ECL2 of CCK-AR interacted with the sulfated Tyr in position-3 of CCK-9. Using site-specific mutagenesis of this Met and CCK-9 analogs, these authors tested this hypothesis and presented evidence that supported an interaction between this Met and Tyr(SO_3H)-3 as being important for ligand binding and receptor activation. In a subsequent study, predictions of the model that the carboxylate of Asp-8 and the carboxyl terminal amide of CCK-9 interacted with Asn and Arg residues at the boundary of TMH6 and ECL3 were tested, and data supporting these interactions were obtained. From these findings, the authors were able to identify a hydrophilic group of receptor residues that interact with CCK-9 and hypothesized that electrostatic interactions were important in converting CCK-AR from an inactive to an active state. Gouldson and co-workers (30) constructed a model of CCK-AR and presented evidence based on mutational analyses that a nonpeptide agonist and antagonist bind via interactions with different residues within the TMH bundle. The antagonist was predicted to bind near the extracellular surface of the TMH bundle via a limited number of contacts whereas the nonpeptide agonist, like CCK-8, bound deeper within the bundle and made many more contacts with receptor residues.

Jarnagin and colleagues (12) constructed a model of the TMH bundle of the B2 bradykinin (BK) receptor based on Br and mutagenesis experiments of several GPCRs. The conformation of BK in solution resolved by nuclear magnetic resonance spectroscopy was docked into the receptor model. Based on predictions of the model, a number of residues in the TMH5, 6, and 7 were mutated and data consistent with interactions between these residues and BK were provided.

Feighner and colleagues (31) constructed an energy-minimized model for the hexapeptide GH-releasing peptide (GHRP-6) receptor. The model predicted interactions of the ligand with residues in TMH2, 3, 5, and 6 and in ECL1 that were tested by site-specific mutations. The data obtained were consistent with the predictions.

Macdonald and colleagues (32) constructed a melanin-concentrating hormone (MCH)/MCH receptor complex by docking MCH in a conformation obtained by molecular dynamics simulation into a receptor model based on a Br template. The model predicted an interaction between an Asp in TMH3 and Arg-11 in the 19-amino acid, cyclic peptide. The prediction was tested experimentally by mutation of the receptor and synthesis of position-11 MCH analogs. The findings were consistent with a role for the Asp in TMH3 in binding MCH and in receptor activation.

Mouillac, Barberis, Hibert, and co-workers (8, 9, 33, 34) have used models of the vasopressin V_{1a} receptor to study binding of agonists and antagonists. Their model predicted that a primary binding pocket for the cyclic, nonapeptide vasopressin was within the TMH bundle (see above), and this was confirmed by mutational analysis of the receptor. They next studied the binding of linear peptide and nonpeptide antagonists. They found that aromatic residues in TMH6 of the receptor were needed for binding antagonists even though they are not involved in binding agonists. These findings provide new insights that could be used in rational design of antagonists *vs.* agonists.

Role of highly conserved Asn in TMH1, Asp in TMH2 and Asn in TMH7 in GPCR structure

Sealfon, Weinstein, and colleagues (35, 36) used models of the GnRH receptor to predict structural features that were important for receptor function. They noticed that the highly conserved pattern of Asn in TMH1, Asp in TMH2 and Asn in TMH7 was changed in GnRH receptor to Asn in TMH1, Asn in TMH2 and Asp in TMH7. A model of the TMH bundle of GnRH receptor was constructed and energy minimized. The model predicted that these three residues interact to stabilize the helical bundle. The findings from experiments using site-specific mutations of these residues and functional analyses of the mutant receptors were consistent with this prediction.

Our model of the unoccupied TRH receptor (see above) predicted that Asn-43, Asp-71and Asn-316 interacted via H-bonds to hold TMH1, 2 and 7 together (37). This prediction was tested by site-specific mutagenesis, and the data obtained were consistent with this hypothesis.

In the 3D crystal structure of Rh (1), the conserved Asn in TMH1 and the Asp in TMH2 interact with the peptide backbone carbonyl one helical turn above the conserved Asn in TMH7, and Asp in TMH2 and Asn in TMH7 interact via a water molecule also. Thus, models provided the first insights into how residues in TMH1, 2, and 7 through intramolecular interactions constrained the conformation of TMH bundle of GPCRs.

Role of highly conserved Asn-Pro/Asp-Pro motif in the structure of TMH7

Weinstein and colleagues (38) used results from analysis of a structure database search and dynamic simulations to construct a model of TMH7 of a GPCR. Their model predicted that TMH7 is not an ideal α-helix but contains a kink caused by the Asn-Pro/Asp-Pro motif. Using the serotonin receptor type 2a as a prototype, they suggested that a model incorporating this kinked helix explained previously generated experimental data. However, the Rh structure shows that the Asn-Pro-Xaa-Xaa-Tyr sequence has a regular helical structure, and a kink occurs at a point further along the helix. Other parts of TMH7, in particular the area around the Lys that covalently attaches to retinal and the sequence near the Ala residue whose carbonyl interacts with Asn in TMH1, are distorted.

Role of highly conserved Asp/Glu-Arg-Tyr sequence in GPCR activation

The amino acid sequence Asp/Glu-Arg-Tyr (D/E-R-Y) at the intracellular end of TMH3 has been shown to be important in receptor activation in several GPCRs. It was initially shown that the charged pair of Glu-Arg was needed for

rhodopsin activation because double mutants of these residues failed to activate transducin (39). Scheer, Fanelli, and colleagues (40, 41) constructed models of native α_{1B}-adrenergic receptors and of receptors with mutations of the Asp and Arg residues of the D-R-Y sequence that exhibited constitutive signaling activities. Functional experiments with these receptors showed that mutation of Arg to several other amino acids caused inhibition of stimulated signaling and to Lys increased basal signaling, whereas mutation of Asp to Ala caused a marked increase in basal activity. Their models predicted that the Arg residue was located in a hydrophilic pocket formed by the highly conserved Asn in TMH1, Asp in TMH2 and Asn and Tyr in TMH7 (see above) in the receptors that were not basally active and that this Arg was not in this pocket in basally active α_{1B}-adrenergic receptor mutants. Based on their experimental observations and model predictions, and the finding that activation of Rh is associated with proton uptake (42), these investigators proposed the following hypothesis for activation of the α_{1B}-adrenergic receptor. The inactive receptor is restrained by interactions between Arg of D-R-Y and the hydrophilic residues of the pocket formed between TMH1, 2 and 7. Activation is secondary to protonation of the Asp in D-R-Y that causes the Arg of the D-R-Y sequence to move out of the TMH bundle and changes the orientation of residues in ICL2 and 3 that allow for increased affinity of coupling to G protein-coupled.

Fanelli and colleagues (43) analyzed the Asp-Arg-Cys, rather than Asp-Arg-Tyr in most GPCRs, sequence at the end of TMH3 in the oxytocin receptor by mutation and modeling. Mutation of the Arg to Ala caused this receptor to become constitutively active. Nevertheless, they predicted that the same movement of Arg from a hydrophilic pocket formed by TMH1, 2, and 7 in the oxytocin receptor creates a crevice formed by ICL2 and 3 and the intracellular extension of TMH5 to which a G protein may couple. This hypothesis is similar to the one they proposed for the α_{1B}-adrenergic receptor (see above).

A different hypothesis regarding a protonation-dependent mechanism of activation involving Arg at the intracellular aspect of TMH3 has been proposed for the GnRH receptor. From their models, Ballesteros and colleagues (36) predicted that the Arg of the Asp-Arg-Ser sequence at the end of TMH3 interacts with the adjacent Asp in the inactive state of the GnRH receptor. Activation involves release of the Arg from interacting with Asp by Asp protonation and promotes movement of the Arg into a hydrophilic pocket in the TMH bundle. An Ile one helical turn above the Arg in TMH3 sterically directs the Arg into the TMH hydrophilic pocket. They provided experimental support for this hypothesis by finding that mutation of Arg markedly decreased agonist-stimulated activation whereas receptors with mutations of Asp to Asn or Ile to Ala were less well stimulated than GnRH receptor.

Thus, these models predict that movement of the sidechain of the Arg of the highly conserved E/D-R-Y sequence is involved in receptor activation. In particular, they support a role for the Arg as a switch to convert a GPCR from an inactive to an active state in response to agonist-stimulated receptor protonation. A recent report, however, has provided evidence that the conserved Asp in the D-R-Y sequence is not necessary for protonation-dependent activation of the β_2-adrenergic receptor even though protonation may be involved (44).

Role of transmembrane helix movement in GPCR activation

Experiments with mutant rhodopsins in which double site-specific substitutions by Cys labeled for electron spin resonance spectroscopy demonstrated that movement of TMHs relative to one another were involved in receptor activation (45). In the case of Rh, TMH3 and 6 were shown to exhibit relative motion upon light activation. Movement of TMHs of Rh upon conversion of 11-*cis*-retinal to all *trans*-retinal is predicted from the crystal structure also (1). Gether and co-workers (46) performed experiments using a conformationally sensitive Cys-reactive reagent and found that positions in TMH3 and 6 changed their environments upon binding β_2-adrenergic agonists. Based on their model of the β_2-adrenergic receptor, they concluded that these movements were similar to those proposed in Rh.

Based on our analyses of unoccupied and TRH-occupied TRH receptor models, we (47) predicted relative movement of TMH5 and 6 upon activation of TRH receptor. To provide evidence in support of this hypothesis, we mutated residues in TMH5 and 6 that our model predicted were involved in interhelical aromatic interactions that constrain the receptor in the inactive conformation. Mutation of these residues to amino acids that could not form aromatic interactions should allow these helices to move apart and thereby activate TRH receptor. As predicted, mutation of Phe-199 in TMH5 and of Trp-279 in TMH6 to Ala produced receptors that were constitutively active.

Lin and colleagues (15) constructed a model of the LH/CG receptor and used it along with an analysis of naturally occurring constitutively activating mutations to predict TMH movements that constitute LH/CG receptor activation. They concluded that hydrophobic interactions among residues in the intracellular halves of TMH5 and 6, and polar interactions between residues in TMH6 and 7 constrain LH/CG receptor in an inactive conformation. They suggested that activation involves agonist-induced disruption of these interactions that allows TMH6 movement. Findings inconsistent with some of these predictions, however, have been reported (16). An analysis of a number of constitutively active LH/CG receptor mutants supports the idea that activation is caused by release of intrahelical interactions (48).

Fanelli and colleagues (43) predicted movements of TMH3, 4, 5, and 6 as components of activation of the oxytocin receptor secondary to movement of the Arg sidechain of the E/D-R-Y sequence (see above). Zhang and Weinstein (49) predicted that TMH5 and 6 underwent the greatest motion when agonists bound to serotonin type 2 receptors.

Biologically active conformation of an agonist

Prevailing theory of GPCR activation holds that the receptor can assume many conformations within a spectrum from inactive to fully active and that different agonists will lead to different subsets of receptor conformations. In general, especially for small ligands, conformational changes of the agonist are not considered. However, to understand receptor activation, the conformation of the active pharma-

cophore is critical. We (50) found that the conformation of TRH in our model of the TRH/TRH receptor complex was different from the predominant conformations of TRH in solution. To gain insight into the biologically active TRH conformation, Dr. Kevin Moeller (Washington University, St. Louis, MO) synthesized conformationally restricted analogs of TRH. We characterized two diastereomers of an analog restricted by addition of a methylene bridge between the second and third positions of a TRH analog, pyroGlu-cyclohexylAla-ProNH_2. We found that one diastereomer bound with higher affinity and exhibited higher intrinsic activity than pyroGlu-cyclohexylAla-ProNH_2, whereas the other diastereomer bound with lower affinity and was less active than pyroGlu-cyclohexylAla-ProNH_2. Computer simulations of these analogs predicted that the higher affinity, more active diastereomer assumed a conformation that was similar to TRH in our model of the TRH/TRH receptor complex whereas the lower affinity, less active diastereomer was similar in structure to TRH in solution. Thus, our model of the TRH/TRH receptor complex predicted a biologically active conformation of TRH. These types of predictions may allow for the design of better agonists of GPCRs.

Conclusions

The emergence of the relationship between structure and function in GPCRs has been made possible by the combination of computational modeling and molecular pharmacological experiments. We are only at the beginning of this process, but we have already made progress in understanding the basis for receptor selectivity, *i.e.* how ligands bind to receptors. We are beginning to delineate the molecular details of the consequences of agonist interaction with receptors, *i.e.* the molecular basis for receptor activation. We are also gaining increasing understanding of how mutated receptors become constitutively active and lead to human disease. One of the most exciting recent developments is the determination of the structure of rhodopsin. Hopefully, structures of other GPCRs will follow. However, while the structure of rhodopsin may provide a guideline to the structures of related receptors, it does not delineate the details of the structures of other receptors and, more importantly, it does not elucidate the structural changes that lead to receptor activation. Thus, a combination of theory and experiment will remain the approach of choice not only to discover the basis for receptor function but also to successfully design therapeutic agents to treat human disease.

Acknowledgments

We thank the members of our research groups for their contributions to the work on the TRH receptor. In particular, we acknowledge the contributions of A.-O. Colson, A. Jinsi-Parimoo, L. Laakkonen, J. H. Perlman, A. Rosenhouse-Dantsker, and W. Wang. We also thank K.D. Moeller (Department of Chemistry, WA University, St. Louis, MO) for synthesis of many of the TRH analogs we used.

We apologize to those investigators whose relevant contributions we failed to include in this review.

References

1. **Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M** 2000 Crystal structure of rhodopsin: a G protein-coupled receptor. Science 289:739–745
2. **Luecke H, Schobert B, Richter HT, Cartailler JP, Lanyi JK** 1999 Structure of bacteriorhodopsin at 1.55 Å resolution. J Mol Biol 291:899–911
3. **Röper D, Jacoby E, Krüger P, Engels M, Grötzinger J, Wollmer A, Strassburger W** 1994 Modeling of G-protein coupled receptors with bacteriorhodopsin as a template. A novel approach based on interaction energy differences. J Recept Res 14:167–186
4. **Baldwin JM** 1993 The probable arrangement of the helices in G protein-coupled receptors. EMBO J 12:1693–1703
5. **Osman R, Colson A-O, Perlman JH, Laakkonen LJ, Gershengorn MC** 1999 Mapping binding sites for peptide G protein-coupled receptors: The receptor for thyrotropin-releasing hormone. In: Wess J (ed) Structure/Function of G-Protein Coupled Receptors. John Wiley & Sons, New York
6. **Gigoux V, Escrieut C, Silvente-Poirot S, Maigret B, Gouilleux L, Fehrentz JA, Gully D, Moroder L, Vaysse N, Fourmy D** 1998 Met-195 of the cholecystokinin-A receptor interacts with the sulfated tyrosine of cholecystokinin and is crucial for receptor transition to high affinity state. J Biol Chem 273:14380–14386
7. **Gigoux V, Escrieut C, Fehrentz JA, Poirot S, Maigret B, Moroder L, Gully D, Martinez J, Vaysse N, Fourmy D** 1999 Arginine 336 and asparagine 333 of the human cholecystokinin-A receptor binding site interact with the penultimate aspartic acid and the C-terminal amide of cholecystokinin. J Biol Chem 274:20457–20464
8. **Phalipou S, Seyer R, Cotte N, Breton C, Barberis C, Hibert M, Mouillac B** 1999 Docking of linear peptide antagonists into the human V_{1a} vasopressin receptor - identification of binding domains by photoaffinity labeling. J Biol Chem 274:23316–23327
9. **Cotte N, Balestre MN, Aumelas A, Mahé E, Phalipou S, Morin D, Hibert M, Manning M, Durroux T, Barberis C, Mouillac B** 2000 Conserved aromatic residues in the transmembrane region VI of the V_{1a} vasopressin receptor differentiate agonist vs. antagonist ligand binding. Eur J Biochem 267:4253–4263
10. **Cui JS, Smith RG, Mount GR, Lo JL, Yu JH, Walsh TF, Singh SB, DeVita RJ, Goulet MT, Schaeffer JM, Cheng K** 2000 Identification of Phe^{313} of the gonadotropin-releasing hormone (GnRH) receptor as a site critical for the binding of nonpeptide GnRH antagonists. Mol Endocrinol 14:671–681
11. **Hoffmann SH, Ter Laak T, Kühne R, Reiländer H, Beckers T** 2000 Residues within transmembrane helices 2 and 5 of the human gonadotropin-releasing hormone receptor contribute to agonist and antagonist binding. Mol Endocrinol 14:1099–1115
12. **Jarnagin K, Bhakta S, Zuppan P, Yee C, Ho T, Phan T, Tahilramani R, Pease JHB, Miller A, Freedman R** 1996 Mutations in the B_2 bradykinin receptor reveal a different pattern of contacts for peptidic agonists and peptidic antagonists. J Biol Chem 271:28277–28286
13. **Lee B, Sharron M, Blanpain C, Doranz BJ, Vakili J, Setoh P, Berg E, Liu G, Guy HR, Durell SR, Parmentier M, Chang CN, Price K, Tsang M, Doms RW** 1999 Epitope mapping of CCR5 reveals multiple conformational states and distinct but overlapping structures involved in chemokine and coreceptor function. J Biol Chem 274:9617–9626
14. **Zhou NM, Luo ZW, Hall JW, Luo JS, Han XB, Huang ZW** 2000 Molecular modeling and site-directed mutagenesis of CCR5 reveal residues critical for chemokine binding and signal transduction. Eur J Immunol 30:164–173
15. **Lin ZL, Shenker A, Pearlstein R** 1997 A model of the lutropin/choriogonadotropin receptor: insights into the structural and functional effects of constitutively activating mutations. Protein Eng 10:501–510
16. **Angelova K, Narayan P, Simon JP, Puett D** 2000 Functional role of transmembrane helix 7 in the activation of the heptahelical lutropin receptor. Mol Endocrinol 14:459–471
17. **Moyle WR, Campbell RK, Rao SNV, Ayad NG, Bernard MP, Han Y, Wang YH** 1995 Model of human chorionic gonadotropin and lutropin receptor interaction that explains signal transduction of the glycoprotein hormones. J Biol Chem 270:20020–20031
18. **Perlman JH, Laakkonen L, Osman R, Gershengorn MC** 1994 A model of the thyrotropin releasing hormone (TRH) receptor binding pocket. Evidence for a second direct interaction between transmembrane helix 3 and TRH. J Biol Chem 269:23383–23386
19. **Perlman JH, Laakkonen L, Osman R, Gershengorn MC** 1995 Distinct roles for arginines in transmembrane helices 6 and 7 of the thyrotropin-releasing hormone receptor. Mol Pharmacol 47:480–484
20. **Perlman JH, Laakkonen LJ, Guarnieri F, Osman R, Gershengorn MC** 1996 A refined model of the thyrotropin-releasing hormone (TRH) receptor binding pocket. Experimental analysis and energy minimization of the complex between TRH and TRH receptor. Biochemistry 35:7643–7650
21. **Laakkonen L, Guarnieri F, Perlman JH, Gershengorn MC, Osman R** 1996 A refined model of the thyrotropin-releasing hormone (TRH) binding pocket. Novel mixed mode MonteCarlo/stochastic dynamics simulations of the complex between TRH and TRH receptor. Biochemistry 35:7651–7663
22. **Ter Laak AM, Timmerman H, Leurs R, Nederkoorn PH, Smit MJ, Donne-Op den Kelder GM** 1995 Modelling and mutation studies on the histamine H1-receptor agonist binding site reveal different binding modes for H1-agonists: Asp116 (TM3) has a constitutive role in receptor stimulation. J Comput Aided Mol Des 9:319–330
23. **Wieland K, Ter Laak AM, Smit MJ, Kühne R, Timmerman H, Leurs R** 1999

Mutational analysis of the antagonist-binding site of the histamine H_1 receptor. J Biol Chem 274:29994–30000
24. Befort K, Tabbara L, Kling D, Maigret B, Kieffer BL 1996 Role of aromatic transmembrane residues of the δ-opioid receptor in ligand recognition. J Biol Chem 271:10161–10168
25. Almaula N, Ebersole BJ, Zhang D, Weinstein H, Sealfon SC 1996 Mapping the binding site pocket of the serotonin 5-Hydroxytryptamine2A receptor. Ser3.36(159) provides a second interaction site for the protonated amine of serotonin but not of lysergic acid diethylamide or bufotenin. J Biol Chem 271:14672–14675
26. Colson AO, Perlman JH, Smolyar A, Gershengorn MC, Osman R 1998 Static and dynamic roles of extracellular loops in G-protein-coupled receptors: a mechanism for sequential binding of thyrotropin-releasing hormone to its receptor. Biophys J 74:1087–1100
27. Perlman JH, Colson A-O, Jain R, Czyzewski B, Cohen LA, Osman R, Gershengorn MC 1997 Role of the extracellular loops of the thyrotropin-releasing hormone receptor: evidence for an initial interaction with thyrotropin-releasing hormone. Biochemistry 36:15670–15676
28. Moro S, Hoffmann C, Jacobson KA 1999 Role of the extracellular loops of G protein-coupled receptors in ligand recognition: a molecular modeling study of the human $P2Y_1$ receptor. Biochemistry 38:3498–3507
29. Kennedy K, Gigoux V, Escrieut C, Maigret B, Martinez J, Moroder L, Fréhel D, Gully D, Vaysse N, Fourmy D 1997 Identification of two amine acids of the human cholecystokinin-A receptor that interact with the N-terminal moiety of cholecystokinin. J Biol Chem 272:2920–2926
30. Gouldson P, Legoux P, Carillon C, Delpech B, Le Fur G, Ferrara P, Shire D 2000 The agonist SR 146131 and the antagonist SR 27897 occupy different sites on the human CCK_1 receptor. Eur J Pharmacol 400:185–194
31. Feighner SD, Howard AD, Prendergast K, Palyha OC, Hreniuk DL, Nargund R, Underwood D, Tata JR, Dean DC, Tan CP, McKee KK, Woods JW, Patchett AA, Smith RG, Van der Ploeg LH 1998 Structural requirements for the activation of the human growth hormone secretagogue receptor by peptide and nonpeptide secretagogues. Mol Endocrinol 12:137–145
32. Macdonald D, Murgolo N, Zhang RM, Durkin JP, Yao XR, Strader CD, Graziano MP 2000 Molecular characterization of the melanin-concentrating hormone/receptor complex: identification of critical residues involved in binding and activation. Mol Pharmacol 58:217–225
33. Chini B, Mouillac B, Ala Y, Balestre MN, Trumpp-Kallmeyer S, Hoflack J, Elands J, Hibert M, Manning M, Jard S, Barberis C 1995 Tyr115 is the key residue for determining agonist selectivity in the V1a vasopressin receptor. EMBO J 14:2176–2182
34. Mouillac B, Chini B, Balestre M-N, Elands J, Trumpp-Kallmeyer S, Hoflack J, Hibert M, Jard S, Barberis C 1995 The binding site of neuropeptide vasopressin V1a receptor. Evidence for a major localization within transmembrane regions. J Biol Chem 270:25771–25777
35. Zhou W, Flanagan C, Ballesteros JA, Konvicka K, Davidson JS, Weinstein H, Millar RP, Sealfon SC 1994 A reciprocal mutation supports helix 2 and helix 7 proximity in the gonadotropin-releasing hormone receptor. Mol Pharmacol 45:165–170
36. Ballesteros J, Kitanovic S, Guarnieri F, Davies P, Fromme BJ, Konvicka K, Chi L, Millar RP, Davidson JS, Weinstein H, Sealfon SC 1998 Functional microdomains in g-protein-coupled receptors - the conserved arginine-cage motif in the gonadotropin-releasing hormone receptor. J Biol Chem 273:10445–10453
37. Perlman JH, Colson A-O, Wang W, Bence K, Osman R, Gershengorn MC 1997 Interactions between conserved residues in transmembrane helices 1, 2 and 7 of the thyrotropin-releasing hormone receptor. J Biol Chem 272:11937–11942
38. Konvicka K, Guarnieri F, Ballesteros JA, Weinstein H 1998 A proposed structure for transmembrane segment 7 of G protein-coupled receptors incorporating an Asn-Pro/Asp-Pro motif. Biophys J 75:601–611
39. Franke RR, Sakmar TP, Graham RM, Khorana HG 1992 Structure and function in rhodopsin. Studies of the interaction between the rhodopsin cytoplasmic domain and transducin. J Biol Chem 267:14767–14774
40. Scheer A, Fanelli F, Costa T, De Benedetti PG, Cotecchia S 1996 Constitutively active mutants of the α_{1B}-adrenergic receptor: role of highly conserved polar amino acids in receptor activation. EMBO J 15:3566–3578
41. Scheer A, Costa T, Fanelli F, De Benedetti PG, Mhaouty-Kodja S, Abuin L, Nenniger-Tosato M, Cotecchia S 2000 Mutational analysis of the highly conserved arginine within the Glu/Asp-Arg-Tyr motif of the α_{1b}-adrenergic receptor: effects on receptor isomerization and activation. Mol Pharmacol 57:219–231
42. Arnis S, Fahmy K, Hofmann KP, Sakmar TP 1994 A conserved carboxylic acid group mediates light-dependent proton uptake and signaling by rhodopsin. J Biol Chem 269:23879–23881
43. Fanelli F, Barbier P, Zanchetta D, De Benedetti PG, Chini B 1999 Activation mechanism of human oxytocin receptor: a combined study of experimental and computer-simulated mutagenesis. Mol Pharmacol 56:214–225
44. Ghanouni P, Schambye H, Seifert R, Lee TW, Rasmussen SGF, Gether U, Kobilka BK 2000 The effect of pH on β_2 adrenoceptor function - evidence for protonation-dependent activation. J Biol Chem 275:3121–3127
45. Farrens DL, Altenbach C, Yang K, Hubbell WL, Khorana HG 1996 Requirement of rigid body motion of transmembrane helices for light activation of rhodopsin. Science 274:768–770
46. Gether U, Lin S, Ghanouni P, Ballesteros JA, Weinstein H, Kobilka BK 1997 Agonists induce conformational changes in transmembrane domains III and VI of the β_2 adrenoceptor. EMBO J 16:6737–6747
47. Colson A-O, Perlman JH, Jinsi-Parimoo A, Nussenzveig DR, Osman R, Gershengorn MC 1998 A hydrophobic cluster between transmembrane helices 5 and 6 constrains the thyrotropin-releasing hormone receptor in an inactive conformation. Mol Pharmacol 54:968–978
48. Fanelli F 2000 Theoretical study on mutation-induced activation of the luteinizing hormone receptor. J Mol Biol 296:1333–1351
49. Zhang D, Weinstein H 1993 Signal transduction by a $5\text{-}HT_2$ receptor: a mechanistic hypothesis from molecular dynamics simulations of the three-dimensional model of the receptor complexed to ligands. J Med Chem 36:934–938
50. Laakkonen L, Li WH, Perlman JH, Guarnieri F, Osman R, Moeller KD, Gershengorn MC 1996 Restricted analogues provide evidence of a biologically active conformation of thyrotropin-releasing hormone. Mol Pharmacol 49:1092–1096
51. Rivkees SA, Barbhaiya H, Ijzerman AP 1999 Identification of the adenine binding site of the human A1 adenosine receptor. J Biol Chem 274:3617–3621
52. Kim J, Wess J, Van Rhee AM, Schöneberg T, Jacobson KA 1995 Site-directed mutagenesis identifies residues involved in ligand recognition in the human A_{2a} adenosine receptor. J Biol Chem 270:13987–13997
53. Kontoyianni M, DeWeese C, Penzotti JE, Lybrand TP 1996 Three-dimensional models for agonist and antagonist complexes with β_2 adrenergic receptor. J Med Chem 39:4406–4420
54. Bramblett RD, Panu AM, Ballesteros JA, Reggio PH 1995 Construction of a 3D model of the cannabinoid CB1 receptor: determination of helix ends and helix orientation. Life Sci 56:1971–1982
55. Javitch JA, Ballesteros JA, Chen JY, Chiappa V, Simpson MM 1999 Electrostatic and aromatic microdomains within the binding-site crevice of the D2 receptor: contributions of the second membrane-spanning segment. Biochemistry 38:7961–7968
56. Mills JS, Miettinen HM, Barnidge D, Vlases MJ, Wimer-Mackin S, Dratz EA, Sunner J, Jesaitis AJ 1998 Identification of a ligand binding site in the human neutrophil formyl peptide receptor using a site-specific fluorescent photoaffinity label and mass spectrometry. J Biol Chem 273:10428–10435
57. Burstein ES, Spalding TA, Brann MR 1998 Structure/function relationships of a G-protein coupling pocket formed by the third intracellular loop of the m5 muscarinic receptor. Biochemistry 37:4052–4058
58. Pepin MC, Yue SY, Roberts E, Wahlestedt C, Walker P 1997 Novel "restoration of function" mutagenesis strategy to identify amino acids of the delta-opioid receptor involved in ligand binding. J Biol Chem 272:9260–9267
59. Ishii I, Izumi T, Tsukamoto H, Umeyama H, Ui M, Shimizu T 1997 Alanine exchanges of polar amino acids in the transmembrane domains of a platelet-activating factor receptor generate both constitutively active and inactive mutants. J Biol Chem 272:7846–7854
60. Moro S, Guo DP, Camaioni E, Boyer JL, Harden TK, Jacobson KA 1998 Human $P2Y_1$ receptor: molecular modeling and site-directed mutagenesis as tools to identify agonist and antagonist recognition sites. J Med Chem 41:1456–1466
61. Underwood DJ, Strader CD, Rivero R, Patchett AA, Greenlee W, Prendergast K 1994 Structural model of antagonist and agonist binding to the angiotensin II, AT1 subtype, G protein coupled receptor. Chem Biol 1:211–221
62. Noda K, Feng YH, Liu XP, Saad Y, Husain A, Karnik SS 1996 The active state of the AT1 angiotensin receptor is generated by angiotensin II induction. Biochemistry 35:16435–16442
63. Prado GN, Mierke DF, Pellegrini M, Taylor L, Polgar P 1998 Motif mutation of bradykinin B2 receptor second intracellular loop and proximal C terminus is critical for signal transduction, internalization, and resensitization. J Biol Chem 273:33548–33555
64. Webb ML, Patel PS, Rose PM, Liu ECK, Stein PD, Barrish J, Lach DA, Stouch T, Fisher SM, Hadjilambris O, Lee H, Skwish S, Dickinson EKJ, Krystek Jr SR 1996 Mutational analysis of the endothelin type A receptor (ET_A): interactions and model of the selective ET_A antagonist EMS-182874 with the putative ET_A receptor binding cavity. Biochemistry 35:2548–2556
65. Prusis P, Frändberg P-A, Muceniece R, Kalvinsh I, Wikberg JES 1995 A three dimensional model for the interaction of MSH with the melanocortin-1 receptor. Biochem Biophys Res Commun 210:205–210
66. Sautel M, Rudolf K, Wittneben H, Herzog H, Martinez R, Munoz M, Eberlein W, Engel W, Walker P, Beck-Sickinger AG 1996 Neuropeptide Y and the nonpeptide antagonist BIBP 3226 share an overlapping binding site at the human Y_1 receptor. Mol Pharmacol 50:285–292
67. Barroso S, Richard F, Nicolas-Ethève D, Reversat JL, Bernassau JM, Kitabgi P, Labbé-Jullié C 2000 Identification of residues involved in neurotensin binding and modeling of the agonist binding site in neurotensin receptor 1. J Biol Chem 275:328–336
68. Giolitti A, Cucchi P, Renzetti AR, Rotondaro L, Zappitelli S, Maggi CA 2000 Molecular determinants of peptide and nonpeptide NK-2 receptor antagonists binding sites of the human tachykinin NK-2 receptor by site-directed mutagenesis. Neuropharmacology 39:1422–1429
69. Ryu K, Gilchrist RL, Tung CS, Ji I, Ji TH 1998 High affinity hormone binding to the extracellular N-terminal exodomain of the follicle-stimulating hormone receptor is critically modulated by exoloop 3. J Biol Chem 273:28953–28958

EXHIBIT E

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 273, No. 28, Issue of July 10, pp. 17299–17302, 1998

Minireview

G Protein-coupled Receptors

I. DIVERSITY OF RECEPTOR-LIGAND INTERACTIONS*

Tae H. Ji‡§, Mathis Grossmann¶, and Inhae Ji‡

From the ‡Department of Molecular Biology, University of Wyoming, Laramie, Wyoming 82071-3944 and ¶Walter and Eliza Hall Institute, Melbourne, Australia

Nearly 2000 G protein-coupled receptors (GPCRs)[1] have been reported since bovine opsin was cloned in 1983 (1) and the β-adrenergic receptor in 1986 (2). They are classified into over 100 subfamilies according to the sequence homology, ligand structure, and receptor function. A substantial degree of amino acid homology is found among members of a particular subfamily, but comparisons between subfamilies show significantly less or no similarity.

Mutations have been observed that relate to a wide spectrum of hereditary and somatic disorders and diseases from cancer to infertility. These mutant receptors are incapable of binding ligand or generating normal signals, constitutively generate signals, or are not appropriately expressed on the cell surface. On the other hand, some mutations are beneficial. For example, a mutation in a chemokine (CCR5) receptor, which is a co-receptor for human immunodeficiency virus (HIV), prevents binding of HIV to target cells and consequently prevents HIV viral infection among the majority of homozygotes with this mutation (3).

Although the majority of GPCRs mediate signal transduction via G proteins, emerging evidence indicates that some of these receptors are also capable of sending signals via alternative signal molecules, *e.g.* Jak2 kinase, phospholipase Cγ, or protein kinase C. These alternative pathways are an indication of the overall diversity occurring in the GPCR superfamily. Furthermore, there are putative seven transmembrane molecules, which do not appear to be coupled to a G protein.

The most striking difference has been observed in the sites and modes of ligand binding and signal generation, which not only manifests the diversity but also indicates the availability of numerous alternative approaches to clinical and industrial applications. In this review, we describe the general structure and ligand interactions of the receptors, and in the following review, Kobilka's group (4) focuses on the conformational changes during receptor activation.

General Structure: N-terminal Segment, Seven TMs, Three Exoloops, Three–Four Cytoloops, and C-terminal Segment

As shown in Fig. 1*A*, all GPCRs have an extracellular N-terminal segment, seven TMs, which form the TM core, three exoloops, three cytoloops, and a C-terminal segment. A fourth cytoplasmic loop is formed when the C-terminal segment is palmitoylated at Cys. Each of the seven TMs is generally composed of 20–27 amino acids. On the other hand, N-terminal segments (7–595 amino acids), loops (5–230 amino acids), and C-terminal segments (12–359 amino acids) vary in size, an indication of their diverse structures and functions. Interestingly, there is a weak positive correlation between an N-terminal segment's length and ligand size (5), suggesting a role in ligand binding, in particular for large polypeptides and glycoprotein hormones. A notable exception is the ~600-amino acid N-terminal segment of neurotransmitter receptors such as the calcium receptor.

Why Seven TMs?—The ubiquitous adoption of a seven TM structure raises the inevitable question concerning its structural and functional merits. Odd numbers of TMs place the N- and C-terminal segments at opposite membrane surfaces. It allows glycosylation and ligand binding at the N-terminal segment, and phosphorylation and palmitoylation at the C-terminal segment for desensitization (6) and internalization. Seven TMs may be the minimum necessary to form six loops and a TM core with a sufficient size and versatility to offer a prodigious number of specificities, regulatory mechanisms, and contact sites for G protein and other signal molecules such as Jak2 kinase (7), phospholipase Cγ (8), GPCR kinases (6), arrestin (6), calmodulin (9), and/or protein kinase C (9). In contrast, five TMs may be insufficient to form a stable yet flexible TM core, whereas nine would be more than enough. For example, ion-gated channel proteins are pentamers of four TM subunits, which form a central hydrophilic channel (5) or a 20-TM monomer (10). In either case, they comprise 20 TMs, and the channel is narrow, excluding ions >5 Å (10). Apparently, four TMs by themselves are not sufficient to form a functional core.

Kinked and Tilted TM α-Helices of Unequal Length and Hydrophobicity—TMs are likely to assume diverse structures. The TMs of bacteriorhodopsin (11) and animal rhodopsin (12) form α-helices although some TM of other membrane proteins such as porins have β structures (13). TM α-helices vary in length and can extend beyond the lipid bilayer. Therefore, the boundaries between TMs and loops are likely to be uneven and may be dynamic. TMs 1, 4, and 7 are significantly more hydrophobic than TMs 2, 3, 5, and 6 that contain several ionic and/or neutral residues. In fact, some of the more hydrophobic TMs 1, 4, and 7 have only one hydrophilic residue such as Asn or Ser. Pro residues are frequently found in TM α-helices, unlike α-helices in globular proteins. A Pro can kink the helix backbone by ~26° and impact the global structure. Some TM α-helices are tilted in the membrane. For example, TM 3 of animal rhodopsin is most tilted by ~30°, whereas TMs 4, 6, and 7 are the least tilted (12). TMs 3, 5, and 6 are suggested to protrude more than the others from the lipid bilayer toward the extracellular surface. TMs often contain Cys residues, and some are believed to be apposed although there is no convincing evidence for a TM disulfide linkage.

Counterclockwise Orientation and Closed Loop Formation of Seven TMs—The seven TMs of bacteriorhodopsin (11), animal rhodopsin (12), and adrenergic receptors (14) are arranged as a closed loop in the counterclockwise direction from TM 1 to TM 7 when viewed from the extracellular surface (Fig. 1*A*). The orientation of the TMs imposes a stereo- and geometric specificity on a ligand's entry into and binding to the TM core. In this arrangement, the core is primarily comprised of TMs 2, 3, 5, 6, and perhaps 7, whereas TMs 1 and 4 are peripherally sequestered. In general, TMs 1, 2, and 7 are apposed (15). This arrangement is consistent with the view that the more hydrophobic TMs 1, 4, and 7 are exposed more to the lipid bilayer than the less hydrophobic TMs 2, 3, 5, and 6. It is unknown whether there are GPCRs that are open-looped or not closed between TMs 1 and 7. Different arrangements can provide interesting mechanisms for ligand binding and receptor activation. For example, if TM 7 associates more closely with TM 2 than TM 1, there will be a long crevice between TMs 1 and 7. Such a crevice offers a binding site for the fatty chain of lysophosphatidic acid. Also, TMs 1 and 7 of a GPCR may associate with TMs 1 and 7 of another GPCR to form a dimer.

TMs are entropically driven into the lipid bilayer primarily by water molecules present outside of the membrane. However, it is unclear what forces are primarily responsible for arranging TMs

* This minireview will be reprinted in the 1998 Minireview Compendium, which will be available in December, 1998. This is the first article of three in the "G Protein-coupled Receptors Minireview Series." This work was supported by Grants HD-18702 and DK-51469 from the National Institutes of Health.

§ To whom correspondence should be addressed. Tel.: 307-766-6272; Fax: 307-766-5098; E-mail: Ji@uwyo.edu.

[1] The abbreviations used are: GPCR, G protein-coupled receptor; HIV, human immunodeficiency virus; TM, transmembrane domain; N-terminal segment, extracellular N-terminal domain; exoloop, extracellular loop; cytoloop, cytoplasmic loop; C-terminal segment, cytoplasmic C-terminal segment; TSH, thyroid-stimulating hormone; PTH, parathyroid hormone; LH, luteinizing hormone; FSH, follicle-stimulating hormone; CG, choriogonadotropin; GABA, γ-aminobutyric acid.

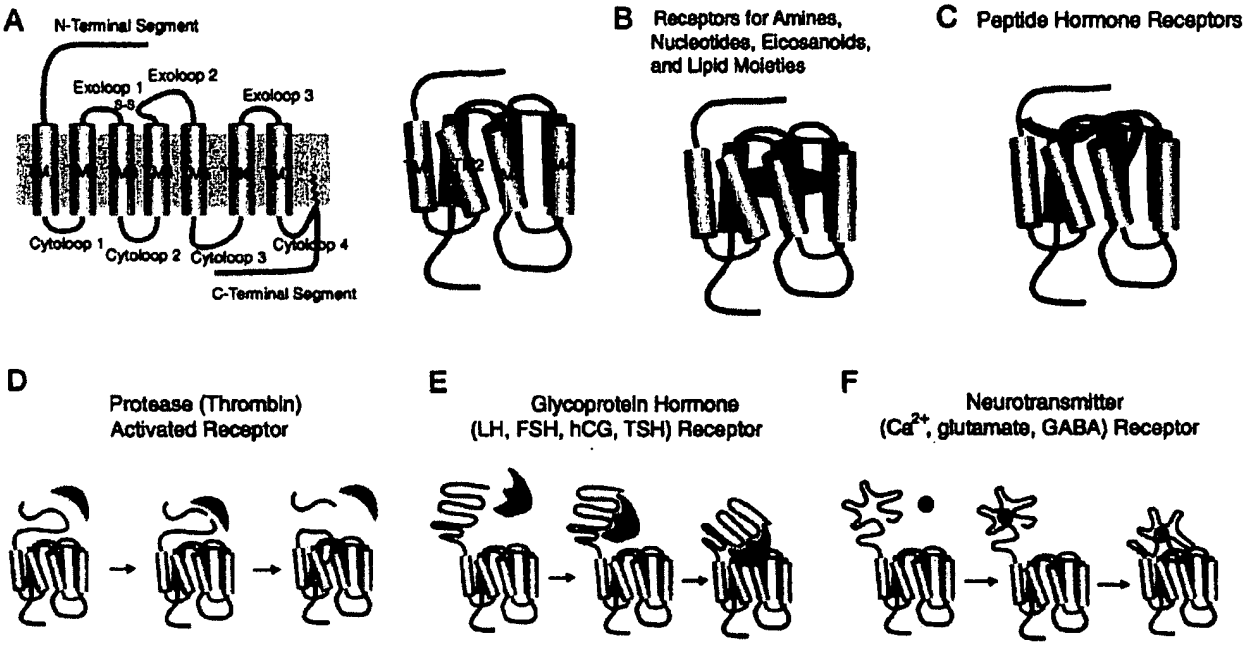

FIG. 1. **Schematic presentation of the general structure of GPCRs and receptor-ligand interactions.** *A*, general structure and terminology. Several distinct modes have been observed for ligand binding and signal generation at exclusively the TM core for photon, biogenic amines, nucleosides, eicosanoids, and moieties (lysophosphatidic acid and sphingosine 1-phosphate) of lipids (*B*), and the core, exoloops, and N-terminal segment for peptides of ≤40 amino acids (*C*). *D*, protease ligands like thrombin bind to and cleave the N-terminal segment. The resulting shorter N-terminal segment interacts with exoloops to generate a signal, whereas the released peptide binds to platelet to stimulate platelet aggregation. *E*, glycoprotein hormones, LH, FSH, human CG, and TSH, bind to the ~350-amino acid N-terminal segment and the liganded N-terminal segment interacts with exoloops to generate a signal. *F*, small neurotransmitters, Ca^{2+}, glutamate, and GABA, bind to the ~600-amino acid N-terminal segment, and the liganded N-terminal segment interacts with the membrane-associated domain, thus generating a signal.

within the lipid bilayer: for example, whether the hydrophilic sides of TMs are entropically driven away from the lipid bilayer, the hydrophobic sides of TMs prefer to associate with the lipid bilayer, or groups in the TM core bond together, *e.g.* by van der Waals interactions.

TM Core Tightly Packed by Hydrogen Bonds and Salt Bridges—The recent high resolution three-dimensional structure of bacteriorhodopsin (11) suggests that the TM core does not have a vacant pocket, channel, or tunnel structure, as generally thought. Although the TM core has polar residues and, therefore, is considered hydrophilic, there are only a limited number (~10) of associated water molecules (11). Extensive hydrogen bonds exist between residues of the same TM as well as other TMs like TMs 2 and 7 (11, 15). In addition to hydrogen bonds, conserved TM 2 Asp pairs with TM 1 and TM 7 Asn residues in the TSH-releasing hormone receptor, and conversely, TM 2 Asn with TM 7 Asp in the GnRH receptor (15). Although TMs have Cys residues and some could be paired, no specific pairs have been identified. Because of these interactions, there may not be space for a TM pore in most GPCRs. The core of bacteriorhodopsin is so tightly packed in the basal state that transfer of a proton or water molecule is unlikely (11). This is reminiscent of the observation that the 20-TM ion channels are not large enough for ions >5 Å to pass (10).

Disulfide Bonds—Disulfides always play a major role in the structure and stability of proteins. Among numerous Cys residues of eukaryotic GPCRs, the two conserved Cys residues in exoloops 1 and 2 are known to be linked by a disulfide in bovine rhodopsin, the TSH-releasing hormone receptor, the thromboxane receptor, and the GnRH receptor. This disulfide is expected to constrain the loops and receptor, specifically pulling exoloop 2 over and thus blocking the extracellular opening of the core. Surprisingly, such a structure is not seen in bacteriorhodopsin, which has no Cys. Instead, exoloop 2 of bacteriorhodopsin is projected away from the core opening and is not near exoloop 1. Furthermore, the two corresponding Cys residues in some GPCRs have been suggested to pair with another Cys rather than with each other (16). Disulfides exist in the N-terminal segment of the glycoprotein hormone receptors (17).

Ligand Binding and Receptor Activation

Several distinct modes (Fig. 1, *B–F*) have been observed for high affinity ligand binding to the TM core exclusively (photon, biogenic amines, nucleosides, eicosanoids, and moieties (lysophosphatidic acid and sphingosine 1-phosphate) of lipids), to both the core and exoloops (peptides ≤40 amino acids), to exoloops and N-terminal segment (polypeptides ≤90 amino acids), or exclusively to the N-terminal segment (glycoproteins ≥30 kDa). The distinction between ligand binding and receptor activation is supported by the existence of antagonists that competitively inhibit agonist binding. Further, glycoprotein hormones bind to the N-terminal segment whereas receptors are activated at the membrane-associated domain (5). Interactions of ligands and receptors seem to involve hydrogen bonds, ion pairs, and hydrophobic contacts.

Ligand Binding and Signal Generation in the TM Core

To define better the mechanics of receptor activation it may be divided into at least three steps, signal generation, TM signal transduction, and signal transfer to cytoplasmic signal molecules (5). The simplest way to activate receptors is for a ligand to bind, generate a signal in, and transduce it through the TM core (Fig. 1*B*). In such a case, changes in TM hydrogen bonds are likely to be the underlying mechanism for signal generation and TM reorganization. For example, photoactivation of retinal straightens its bent configuration and rearranges hydrogen bonds and the packing of TMs 3 and 5–7. In biogenic amine receptors, TM 3 is believed to be the primary site for ligand binding whereas TMs 5 and 6 are the signal generation site. Although initial ligand binding and signal generation are related yet separate phenomena, it is technically difficult to distinguish them. However, a few receptor systems such as glycoprotein hormone receptors and protease-activated receptors are available as good models to distinguish experimentally ligand binding from signal generation and thus investigate the transition between these two steps.

Rhodopsin—The family of photoreceptors comprises several isoforms specific for red, blue, and green colors. These color characteristics stem from variations in their amino acid sequences. All of them have a retinal chromophore, which is attached through for-

mation of a Schiff's base between the aldehyde moiety of retinal and the ϵ-amine of the Lys in the middle of TM 7 (18). The protonated Schiff's base is paired with conserved Glu^{113} present at the boundary between TM 3 and exoloop 1, thus bringing TMs 3 and 7 in apposition. Photoaffinity labeling, mutational analysis, and modeling show that the β-ionone ring of retinal associates with TMs 3, 5, and 6, particularly TM 6 Trp^{265} and Tyr^{268} (19). The C^9 methyl group of 11-*cis*-retinal appears to associate with TM 3 Gly^{121} (19). Light absorption causes a bent 11-*cis* to linear all-*trans* isomerization and part of the TM peptide backbone to be exposed. As a result at least one water hydrogen bond is formed (20) and the salt bridge constraint between TM 3 Glu^{113} and TM 7 Lys^{296} may break. Because each hydrogen bond energy is worth 3–7 kcal/mol, these changes are likely and sufficient to rearrange TMs, particularly TMs 3, 6 (19), and 7, thereby generating a signal.

Biogenic Amine Receptors—Biogenic amines, epinephrine, norepinephrine, dopamine, and histamine enter and bind the TM core (4, 18). The amine of catecholamines pairs with the carboxyl group of TM 3 Asp^{113} of the receptor, 9 Å deep in the core. TM 3 Asp^{113} is crucial for ligand binding but not for signal generation. The fact that Asp^{113} could not be replaced with Glu underscores the tight space and importance of the size and orientation of the carboxyl chain. The catechol ring of the ligand is thought to dock the pocket consisting of TMs 5 and 6. Specifically, the *meta* and *para* OH groups of the catechol ring appear to hydrogen bond, respectively, to TM 5 Ser^{204} and Ser^{207}, which lie one α-helical turn apart on the same side of the TM 5 α-helix. Because this double hydrogen bonding would constrain the TM 5 α-helix (18) and/or the ligand and the ligand-linked TM 3 α-helix, the TM structure and packing are expected to be adjusted. In fact, the TM 6 structure, particularly Cys^{285}, changes in activated receptors (21), and Phe^{290} is thought to stabilize the catechol ring (18). Because the interactions of the catechol ring are more important for receptor activation than ligand binding, these changes could generate a signal. In addition, the interaction of TM 3 Asp^{125} and TM 7 Lys^{331} may break upon receptor activation (22). For histamine, the amine may ion pair with the TM 3 Asp, whereas the imidazole ring associates with TM 5 Asp and Thr, which correspond to TM 5 Ser^{204} and Ser^{207} of catecholamine receptors (18). Acetylcholine also appears to bind similarly the TM core, particularly TMs 5 and 6 for signal generation (18).

Nucleoside and Nucleotide Receptors—Extracellular adenosine, adenine nucleotides, and urdine nucleotides are capable of binding G protein-coupled metabotropic receptors (P2Y receptors). Studies of metabotropic adenosine receptors suggest that ligands bind primarily in the TM core and exoloops. For the adenosine A_1 receptor ligand binding requires TMs 1–4 (23), 6, and 7 as well as exoloop 2 (24). On the other hand, TMs 5–7, particularly Ser^{277} and His^{278} of TM 7, are important for ligand binding by the adenosine A_2 receptor (25). Signal generation by adenosine receptors appears to require additional unknown ligand contacts, because agonists and antagonists bind the receptors differently. Four subclasses of chemoattractant cAMP receptors are found in *Dictyostelium*. These receptors are coupled to a G protein that activates adenylyl cyclase. For ligand binding by the cAMP receptor 1, TMs 3, 5, and 6 and exoloop 3 are important, and exoloop 2 appears to modulate ligand access to the TM core binding site (26).

Eicosanoid Receptors—Eicosanoids are derived from arachidonic acid and have a carboxyl group. This family of lipid hormones, which includes leukotrienes and prostanoids (prostaglandins, prostacyclins, and thromboxanes), binds specific subfamily receptor members of high homology. The ligands bind the TM core, and interactions with TMs 3, 6, and 7 are crucial (27, 28). A conserved TM 7 Arg near the extracellular surface is thought to pair with the carboxyl group of the ligand (27). The structure and orientation of the cyclopentane ring of prostanoids influence receptor specificity (28), suggesting variations in the TM core structure.

Short Peptide Receptors with Ligand Binding Partially in the TM Core and Exoloops

Formyl Receptor—Tripeptide *N*-formyl-Met-Leu-Phe is one of the smallest peptide ligands for GPCRs. Spectroscopic and mutational studies (50) show that the *N*-formyl moiety of the ligand binds in the TM core around TMs 2 and 3, whereas the C-terminal region of the ligands associates with the N-terminal segment and exoloops 1 and 2 (29). Recently, the Val^{83}-Arg^{84}-Lys^{85} sequence that joins TM2 and exoloop 1 has been shown to be photoaffinity labeled with the ligand (30).

Glycoprotein Hormone-releasing Hormone Receptors—Mutational analyses of the thyrotropin-releasing hormone (pyroEHP-NH_2) receptor show that several amino acids in TM 3 and one in exoloop 2 are important for ligand binding. Mutation and affinity-labeling studies of gonadotropin-releasing hormone suggest the N terminus of GnRH binds the TM core and the C terminus around exoloops 2 and 3. His^2 of GnRH is believed to be close to TM 3 Lys^{121}, Arg^8 of GnRH to exoloop 3 Asp^{302}, and the C-terminal glycinamide to exoloop 2 Asn^{102} (15). Also, the N-terminal segment appears to be near the ligand (31). Because TM 3 Lys^{121} is ~7.5 Å away from the extracellular surface and its side chain could extend to near the surface, GnRH enters, but not deeply into, the TM core.

Angiotensin Receptor—The hydrophobic C-terminal region of angiotensin II (DRVYIHPF) appears to enter the TM core of the receptor, and the C-terminal carboxyl group pairs with TM 5 Lys^{199} (32), 7–14 Å from the extracellular surface. On the other hand, Asp-Arg of the DRVYIHPF sequence seems to ion pair with exoloop 2 His^{183} and exoloop 3 Asp^{281} of the receptor, respectively (33). The interaction with Asp^{281} is more crucial for signal generation.

Multistep Contacts at the N-terminal Segment and Exoloops

Receptors for Glucagon, Calcitonin, and Vasoactive Intestinal Peptides—This class of peptide hormones is 30–40 amino acids long. On the other hand, the N-terminal segments of their receptors are 116–147 amino acids long and primarily responsible, but not sufficient, for high affinity ligand binding, as exoloops also appear to be required (34). Partial entry of the ligand into the TM core is a possibility. Gly^{10}-Arg^{24} of calcitonin is essential for interaction with the N-terminal segment of the receptor. The resulting transient hormone-receptor complex, including particularly the N-terminal 10 residues of calcitonin, may make secondary contact with exoloops and generate a signal (34). In the glucagon receptor, Asp^{126}-Lys^{137} of the 145-amino acid N-terminal segment and exoloop 1 are crucial for ligand binding (35).

Parathyroid Hormone (PTH) Receptors—The C-terminal region of 84-amino acid PTH is important for receptor binding, whereas the N-terminal region is crucial for receptor activation. For the receptor, ligand specificity resides primarily in the N-terminal segment whereas signal is generated in the membrane-associated domain, particularly involving TM 3 Ile^{244} and exoloop 2 Tyr^{318} (36). These results suggest the two-step interaction of PTH and the receptor, first to form the transient PTH-N-terminal segment complex and next for the complex to interact with the membrane-associated domain to generate a signal.

Initial Ligand Binding at the N-terminal Segment and Subsequent Secondary Contact of the N-terminal Segment with Exoloops

Protease-activated Receptors—This family of receptors for thrombin and other proteases is activated by proteolysis of the N-terminal segment (37). The protease thrombin recognizes the sequence L^{38}DPRSFLLRNPNDKYEPF^{55} in the 74-amino acid N-terminal segment of the thrombin receptor and cleaves it at the thrombin cleavage sequence, $L^{38}DPR \downarrow S^{42}$. The N-terminal segment is sufficient for the proteolysis by thrombin, and the released peptide, Met^1-Arg^{41}, is a strong agonist to stimulate platelet aggregation (38). At the same time, the resulting new 33-amino acid N-terminal segment acts as a tethered ligand and intramolecularly binds to exoloops of the remaining receptor; in particular Arg^5 of the new N-terminal segment interacts with exoloop 2 Glu^{260} to generate the signal (39). Protease-activated receptor 2 is structurally similar to thrombin receptors and is cleaved at $R^{36} \downarrow S^{37}$ in the N-terminal segment. The resulting new N-terminal segment contacts exoloops, most likely between the conserved Arg of the N-terminal segment and Glu in exoloop 2 (40).

Glycoprotein Hormone Receptors—The glycoprotein hormone receptors consist of two roughly equal domains, a 350–400-residue N-terminal segment and a ~300-amino acid membrane-associated domain. The N-terminal segment alone is capable of high affinity ligand binding whereas the membrane-associated domain is the

site of receptor activation (5). The N-terminal segment has 8–9 Leu-rich repeats. These Leu-rich repeats are thought to form a crescent with the concave inner surface consisting of β-sheets, which may bind ligands (41, 42). In addition to the Leu-rich crescent, the N-terminal region of the N-terminal segment contacts the hormone (43). Although the N-terminal segment alone is capable of high affinity ligand binding, this is modulated by exoloops of the receptor (44).

Glycoprotein hormones, LH, FSH, CG, and TSH, are the largest (30–40 kDa) and most complex GPCR agonists. They are heterodimers of a common α-subunit and a hormone-specific β-subunit. These hormones initially bind exclusively to the N-terminal segment with high affinity. The hormone-N-terminal segment complex undergoes conformational changes (45) and makes secondary interactions with the membrane-associated domain, thus generating a signal (5). The initial high affinity interaction includes multiple contacts between the N-terminal segment and both subunits of the hormone (41, 42). They involve the α C-terminal region, the *N*-oligosaccharide at αAsn^{52}, and a unique α-helix in the α-subunit, as well as an unusual loop (seat belt) in the β-subunit and the peripheral β-hairpin loops of both subunits (41). Secondary contacts occur between the liganded N-terminal segment and exoloops 1–3 (5, 44, 46). Signal is, therefore, likely generated at exoloops. In fact, some mutations in exoloops 1–3 result in constitutive activation (47).

Neurotransmitter Receptors—This family of receptors is noted for the longest N-terminal segments (~600 amino acids) that recognize glutamate, γ-aminobutyric acid (GABA), and calcium/cations.

For metabotropic glutamate receptors, the N-terminal half of the ~600-amino acid N-terminal segment not only provides the ligand binding site and specificity but also indicates the signal specificity for activation of phospholipase C or inhibition of adenylyl cyclase (48). Some metabotropic glutamate receptors bind calcium as well as glutamate, and Ser^{166} appears to be important (49). The structure and size of the $GABA_B$ and calcium receptor are similar to those of the metabotropic glutamate receptors. Therefore, they are expected to similarly bind ligands and be activated.

Conclusions and Future Directions

An increasing number of GPCR subfamilies shows diverse modes of ligand binding, signal generation, TM signal transduction, and signal transfer to various cytoplasmic signal molecules other than G protein, such as Jak2 kinase (7), phospholipase Cγ (8), and protein kinase C (9). Therefore, the question rises whether these distinct responses are induced by a single signal originating from a single receptor or by separate signals generated by a single receptor or different receptors. If a single receptor is capable of activating one or more effectors, does it activate them simultaneously or only one at a time? The general underlying mechanisms appear to involve changes in interactions and conformation, particularly hydrogen bonds, salt bridges, TM rearrangement, and perhaps receptor oligomerization (4).

Acknowledgments—We thank Drs. Hanlee Ji, Kurt W. Miller, Brian K. Kobilka, Martha Vaughan, Peter Nollert, Ehud M. Landua, and Lee E. Limbird for helpful comments.

REFERENCES

1. Nathans, J., and Hogness, D. S. (1983) *Cell* **34**, 807–814
2. Dixon, R. A., Kobilka, B. K., Strader, D. J., Benovic, J. L., Dohlman, H. G., Frielle, T., Bolanowski, M. A., Bennett, C. D., Rands, E., Diehl, R. E., *et al.* (1986) *Nature* **321**, 75–79
3. Liu, R., Paxton, W. A., Choe, S., Ceradini, D., Martin, S. R., Horuk, R. M., MacDonald, M. E., Stuhlmann, H., Koup, R. A., and Landau, N. R. (1996) *Cell* **86**, 367–377
4. Gether, U., and Kobilka, B. (1998) *J. Biol. Chem.* **273**, 17979–17982
5. Ji, T. H., Murdoch, W. J., and Ji, I. (1995) *Endocrine* **3**, 187–194
6. Lefkowitz, R. (1998) *J. Biol. Chem.* **273**, 18677–18680
7. Ali, M., Sayeski, P., Dirksen, L., Hayzer, D., Marrero, M., and Bernstein, K. (1997) *J. Biol. Chem.* **272**, 23382–23388
8. Marrero, M., Venema, R., Ju, H., Venema, V., Schiefer, B., Harp, J., Ling, B., and Eaton, D. (1998) *J. Biol. Chem.* **273**, 7703–7708
9. Minakami, R., Jinnai, N., and Sugiyama, H. (1997) *J. Biol. Chem.* **272**, 20291–20298
10. Stryer, L. (1995) *Biochemistry*, 4th Ed., pp. 300–302, Freeman and Co., New York
11. Pebay-Peyroula, E., Rummel, G., Rosenbusch, J., and Landau, E. (1997) *Science* **277**, 1676–1681
12. Unger, V., Hargrave, P., Baldwin, J., and Schertler, G. (1997) *Nature* **389**, 203–206
13. Schirmer, T., Keller, T., Wang, Y., and Rosenbusch, J. (1995) *Science* **267**, 512–514
14. Mizobe, T., Mervyn Maze, M., Lam, V., Suryanarayana, S., and Kobilka, B. (1996) *J. Biol. Chem.* **271**, 2387–2389
15. Sealfon, S., Weinstein, H., and Millar, R. (1997) *Endocr. Rev.* **18**, 180–205
16. Couet, J., Sar, S., Jolivet, A., Vu Hai, M.-T., Milgrom, E., and Misrahi, M. (1996) *J. Biol. Chem.* **271**, 4545–4552
17. Zhang, R., Buczko, E., and Dufau, M. (1996) *J. Biol. Chem.* **271**, 5755–5760
18. Strader, C. D., Fong, T. M., Tota, M. R., Underwood, D., and Dixon, R. A. (1994) *Annu. Rev. Biochem.* **63**, 101–132
19. Han, M., Groesbeek, M., Sakmar, T., and Smith, S. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 13442–13447
20. Rath, P., Delange, F., Degrip, W., and Rothschild, K. (1998) *Biochem. J.* **329**, 713–717
21. Javitch, J., Fu, D., Liapakis, G., and Chen, J. (1997) *J. Biol. Chem.* **272**, 18546–18549
22. Porter, J. E., Hwa, J., and Perez, D. M. (1996) *J. Biol. Chem.* **271**, 28318–28323
23. Rivkees, S., Lasbury, M., and Barbhaiya, H. (1995) *J. Biol. Chem.* **270**, 20485–20490
24. Olah, M., Jacobson, K., and Stiles, G. (1994) *J. Biol. Chem.* **269**, 24692–24698
25. Kim, J., Wess, J., van Rhee, A., Schöneberg, T., and Jacobson, K. (1995) *J. Biol. Chem.* **270**, 13987–13997
26. Kim, J.-Y., Caterina, M., Milne, L., Lin, K., Borleis, J., and Devreotes, P. (1997) *J. Biol. Chem.* **272**, 2060–2068
27. Funk, C., Furci, L., Moran, N., and Fitzgerald, G. (1993) *Mol. Pharmacol.* **44**, 934–939
28. Kobayashi, T., Kiriyama, M., Hirata, T., Hirata, M., Ushikubi, F., and Narumiya, S. (1997) *J. Biol. Chem.* **272**, 15154–15160
29. Perez, H., Vilander, L., Andrews, W., and Holmes, R. (1994) *J. Biol. Chem.* **269**, 22485–22487
30. Mills, J. S., Miettinen, H. M., Barnidge, D., Vlases, M. J., Wimer-Mackin, S., Dratz, E. A., Sunner, J., and Jesaitis, A. J. (1998) *J. Biol. Chem.* **273**, 10428–10435
31. Davidson, J., Assefa, D., Pawson, A., Davies, P., Hapgoo, D. J., Becker, I., Flanagan, C., Roeske, R., and Millar, R. (1997) *Biochemistry* **36**, 12881–12889
32. Noda, K., Saad, Y., Kinoshita, A., Boyle, T., Graham, R., Husain, A., and Karnik, S. (1995) *J. Biol. Chem.* **270**, 2284–2289
33. Fen, G. Y., Noda, K., Saad, Y., Liu, X., Husain, A., and Karnik, S. (1995) *J. Biol. Chem.* **270**, 12846–12850
34. Stroop, S., Nakamuta, H., Kuestner, R., Moore, E., and Epand, R. (1996) *Endocrinology* **137**, 4752–4756
35. Unson, C., Cypess, A., Wu, C., Goldsmith, P., Merrifield, R., and Sakmar, T. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 310–315
36. Bergwitz, C., Jusseaume, S., Luck, M., Juppner, H., and Gardella, T. (1997) *J. Biol. Chem.* **272**, 28861–28868
37. Vu, T., Hung, D., Wheaton, V., and Coughlin, S. (1991) *Cell* **64**, 1057–1068
38. Furman, M., Liu, L., Benoit, S., Becker, R., Barnard, M., and Michelson, A. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 3082–3087
39. Nanevicz, T., Wang, L., Chen, M., Ishii, M., and Coughlin, S. (1996) *J. Biol. Chem.* **271**, 702–706
40. Lerner, D. J., Chen, M., Tram, T., and Coughlin, S. R. (1996) *J. Biol. Chem.* **271**, 13943–13947
41. Grossmann, M., Weintraub, B., and Szkudlinski, M. (1997) *Endocr. Rev.* **18**, 476–501
42. Dias, J. A., Lindau-Shepard, B., Hauer, C., and Auger, I. (1998) *Biol. Reprod.* **58**, 1331–1336
43. Phang, T., Kundu, G., Hong, S., Ji, I., and Ji, T. H. (1998) *J. Biol. Chem.* **273**, 13841–13847
44. Ryu, K., Lee, H., Kim, S., Beauchamp, J., Tung, C., Isaacs, N. W., Ji, I., and Ji, T. H. (1998) *J. Biol. Chem.* **273**, 6285–6291
45. Ji, I., Pan, Y.-N., Lee, Y.-M., Phang, T., and Ji, T. H. (1995) *Endocrine* **3**, 907–911
46. Remy, J. J., Couture, L., Pantel, J., Haertle, T., Rabesona, H., Bozon, V., Pajot-Augy, E., Robert, P., Troalen, F., Salesse, R., and Bidart, J. M. (1996) *Mol. Cell. Endocrinol.* **125**, 79–91
47. Parma, J., Van Sande, J., Swillens, S., Tonacchera, M., Dumont, J., and Vassart, G. (1995) *Mol. Endocrinol.* **9**, 725–733
48. Takahashi, K., Tsuchida, K., Tanabe, Y., Masu, M., and Nakanishi, S. (1993) *J. Biol. Chem.* **268**, 19341–19345
49. Kubo, Y., Miyashita, T., and Murata, Y. (1998) *Science* **279**, 1722–1725
50. Fay, S. P., Domalewski, M. D., and Sklar, L. A. (1993) *Biochemistry* **32**, 1627–1631

1: Adv Exp Med Biol. 1998;449:371-85.

Identification of neurohypophysial hormone receptor domains involved in ligand binding and G protein coupling.

Postina R, Kojro E, Fahrenholz F.

Max-Planck-Institut fur Biophysik, Frankfurt am Main, Germany.

Chimeric vasopressin V2/OT receptors were constructed and investigated to identify receptor regions involved in ligand binding or G protein coupling. The fusion sites for one series of hybrid receptors were either located at the C-terminal end of the third extracellular domain or in the centre of the third transmembrane helix, respectively. In each pair of the resulting symmetrical hybrids only one receptor was able to bind arginine vasopressin (AVP) and/or oxytocin (OT). In both cases a major part of the vasopressin V2 receptor (V2R) was needed for ligand binding. A chimeric OT/V2 receptor including OT receptor (OTR) sequences from its N-terminus to the middle of transmembrane region three showed both high-affinity OT binding (Ki = 3 nM) and activation of the adenylyl cyclase. In contrast, a hybrid containing OTR sequences reaching from transmembrane helix five to its C-terminus showed the V2 receptor's ligand binding profile and was unable to couple to G alpha s. These results indicate (i) that the third and/or the fourth intracellular domain of the V2R are involved in G protein coupling and (ii) for high-affinity OT binding the N-terminal third of the OTR plays an important role. By detailed binding studies on a second series of chimeric V2/OT receptors with AVP, OT and the two hybrid hormone derivatives arginine vasotocin and oxypressin it was further demonstrated that the first two extracellular domains of the OTR are involved in binding to the C-terminal tripeptide of OT. Moreover, the third extracellular domain of the OTR is able to contact the cyclic part of OT and the fourth outer domain does not interact with the two variable amino acid residues of AVP and OT. Thus, the first three extracellular domains of the OTR provide an essential part of the OT binding site. The other part is most probably contributed by the OTR's transmembrane helices 3 and 4. Photoaffinity labeling and ligand binding studies demonstrated that the binding site for the OT antagonist d(CH2)5[Tyr(Me)2, Thr4, Orn8, Tyr9]vasotocin is located in the helices 1, 2 and 7. Our results provide evidence for the existence of separate domains of a peptide hormone receptor involved in binding and selectivity for agonists and peptide antagonists.

Publication Types:
- Review
- Review, Tutorial

PMID: 10026828 [PubMed - indexed for MEDLINE]

EXHIBIT G

Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions

JAMES U. BOWIE,* JOHN F. REIDHAAR-OLSON, WENDELL A. LIM, ROBERT T. SAUER

n amino acid sequence encodes a message that deter-ines the shape and function of a protein. This message is ghly degenerate in that many different sequences can de for proteins with essentially the same structure and tivity. Comparison of different sequences with similar essages can reveal key features of the code and improve derstanding of how a protein folds and how it per-rms its function.

THE GENOME IS MANIFEST LARGELY IN THE SET OF PROteins that it encodes. It is the ability of these proteins to fold into unique three-dimensional structures that allows them to ıction and carry out the instructions of the genome. Thus, nprehending the rules that relate amino acid sequence to struc-e is fundamental to an understanding of biological processes. cause an amino acid sequence contains all of the information cessary to determine the structure of a protein (*1*), it should be ssible to predict structure from sequence, and subsequently to er detailed aspects of function from the structure. However, both oblems are extremely complex, and it seems unlikely that either l be solved in an exact manner in the near future. It may be ssible to obtain approximate solutions by using experimental data simplify the problem. In this article, we describe how an analysis allowed amino acid substitutions in proteins can be used to uce the complexity of sequences and reveal important aspects of ıcture and function.

ethods for Studying Tolerance to quence Variation

There are two main approaches to studying the tolerance of an ino acid sequence to change. The first method relies on the cess of evolution, in which mutations are either accepted or cted by natural selection. This method has been extremely verful for proteins such as the globins or cytochromes, for which uences from many different species are known (*2–7*). The second roach uses genetic methods to introduce amino acid changes at specific positions in a cloned gene and uses selections or screens to identify functional sequences. This approach has been used to great advantage for proteins that can be expressed in bacteria or yeast, where the appropriate genetic manipulations are possible (*3, 8–11*). The end results of both methods are lists of active sequences that can be compared and analyzed to identify sequence features that are essential for folding or function. If a particular property of a side chain, such as charge or size, is important at a given position, only side chains that have the required property will be allowed. Conversely, if the chemical identity of the side chain is unimportant, then many different substitutions will be permitted.

Studies in which these methods were used have revealed that proteins are surprisingly tolerant of amino acid substitutions (*2–4, 11*). For example, in studying the effects of approximately 1500 single amino acid substitutions at 142 positions in *lac* repressor, Miller and co-workers found that about one-half of all substitutions were phenotypically silent (*11*). At some positions, many different, nonconservative substitutions were allowed. Such residue positions play little or no role in structure and function. At other positions, no substitutions or only conservative substitutions were allowed. These residues are the most important for *lac* repressor activity.

What roles do invariant and conserved side chains play in proteins? Residues that are directly involved in protein functions such as binding or catalysis will certainly be among the most conserved. For example, replacing the Asp in the catalytic triad of trypsin with Asn results in a 10^4-fold reduction in activity (*12*). A similar loss of activity occurs in λ repressor when a DNA binding residue is changed from Asn to Asp (*13*). To carry out their function, however, these catalytic residues and binding residues must be precisely oriented in three dimensions. Consequently, mutations in residues that are required for structure formation or stability can also have dramatic effects on activity (*10, 14–16*). Hence, many of the residues that are conserved in sets of related sequences play structural roles.

Substitutions at Surface and Buried Positions

In their initial comparisons of the globin sequences, Perutz and co-workers found that most buried residues require nonpolar side chains, whereas few features of surface side chains are generally conserved (*6*). Similar results have been seen for a number of protein families (*2, 4, 5, 7, 17, 18*). An example of the sequence tolerance at surface versus buried sites can be seen in Fig. 1, which shows the allowed substitutions in λ repressor at residue positions that are near the dimer interface but distant from the DNA binding surface of the protein (*9*). These substitutions were identified by a functional

authors are in the Department of Biology, Massachusetts Institute of Technology, bridge, MA 02139.

ıent address: Department of Chemistry and Biochemistry and the Molecular gy Institute, University of California, Los Angeles, Los Angeles, CA 90024.

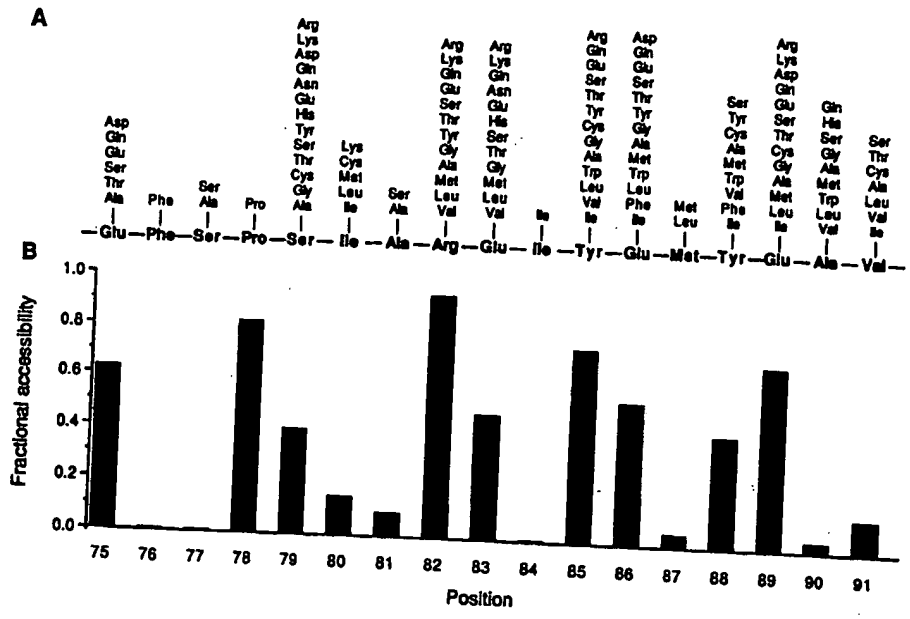

Fig. 1. (**A**) Amino acid substitutions allowed in a short region of λ repressor. The wild-type sequence is shown along the center line. The allowed substitutions shown above each position were identified by randomly mutating one to three codons at a time by using a cassette method and applying a functional selection (*9*). (**B**) The fractional solvent accessibility (*42*) of the wild-type side chain in the protein dimer (*43*) relative to the same atoms in an Ala-X-Ala model tripeptide.

selection after cassette mutagenesis. A histogram of side chain solvent accessibility in the crystal structure of the dimer is also shown in Fig. 1. At six positions, only the wild-type residue or relatively conservative substitutions are allowed. Five of these positions are buried in the protein. In contrast, most of the highly exposed positions tolerate a wide range of chemically different side chains, including hydrophilic and hydrophobic residues. Hence, it seems that most of the structural information in this region of the protein is carried by the residues that are solvent inaccessible.

Constraints on Core Sequences

Because core residue positions appear to be extremely important for protein folding or stability, we must understand the factors that dictate whether a given core sequence will be acceptable. In general, only hydrophobic or neutral residues are tolerated at buried sites in proteins, undoubtedly because of the large favorable contribution of the hydrophobic effect to protein stability (*19*). For example, Fig. 2 shows the results of genetic studies used to investigate the substitutions allowed at residue positions that form the hydrophobic core of the NH_2-terminal domain of λ repressor (*20*). The acceptable core sequences are composed almost exclusively of Ala, Cys, Thr, Val, Ile, Leu, Met, and Phe. The acceptability of many different residues at each core position presumably reflects the fact that the hydrophobic effect, unlike hydrogen bonding, does not depend on specific residue pairings. Although it is possible to imagine a hypothetical core structure that is stabilized exclusively by residues forming hydrogen bonds and salt bridges, such a core would probably be difficult to construct because hydrogen bonds require pairing of donors and acceptors in an exact geometry. Thus the repertoire of possible structures that use a polar core would probably be extremely limited (*21*). Polar and charged residues are occasionally found in the cores of proteins, but only at positions where their hydrogen bonding needs can be satisfied (*22*).

The cores of most proteins are quite closely packed (*23*), but some volume changes are acceptable. In λ repressor, the overall core volume of acceptable sequences can vary by about 10%. Changes at individual sites, however, can be considerably larger. For example, as shown in Fig. 2, both Phe and Ala are allowed at the same core position in the appropriate sequence contexts. Large volume changes at individual buried sites have also been observed in phylogenetic studies, where it has been noted that the size decreases and increases at interacting residues are not necessarily related in a simple complementary fashion (*5, 7, 17*). Rather, local volume changes are accommodated by conformational changes in nearby side chains and by a variety of backbone movements.

The Informational Importance of the Core

With occasional exceptions, the core must remain hydrophobic and maintain a reasonable packing density. However, since the core is composed of side chains that can assume only a limited number of conformations (*24*), efficient packing must be maintained without steric clashes. How important are hydrophobicity, volume, and steric complementarity in determining whether a given sequence can form an acceptable core? Each factor is essential in a physical sense, as a stable core is probably unable to tolerate unsatisfied hydrogen bonding groups, large holes, or steric overlaps (*25*). However, in an informational sense, these factors are not equivalent. For example, in experiments in which three core residues of λ repressor were mutated simultaneously, volume was a relatively unimportant informational constraint because three-quarters of all possible combinations of the 20 naturally occurring amino acids had volumes within the range tolerated in the core, and yet most of these sequences were unacceptable (*20*). In contrast, of the sequences that contained only

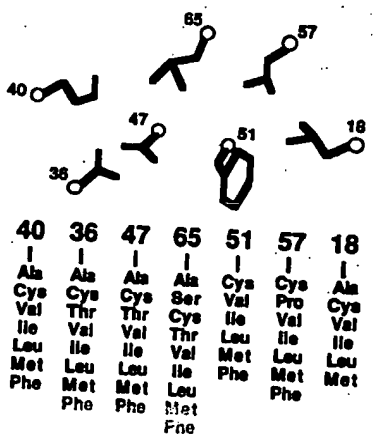

Fig. 2. Amino acid substitutions allowed in the core of λ repressor. The wild-type side chains are shown pictorially in the approximate orientation seen in the crystal structure (*43*). The lists of allowed substitutions at each position are shown below the wild-type side chains. These substitutions were identified by randomly mutating one to four residues at a time by using a cassette method and applying a functional selection (*20*). Not all substitutions are allowed in every sequence background.

the appropriate hydrophobic residues, a significant fraction were acceptable. Hence, the hydrophobicity of a sequence contains more information about its potential acceptability in the core than does the total side chain volume. Steric compatibility was intermediate between volume and hydrophobicity in informational importance.

The Informational Importance of Surface Sites

We have noted that many surface sites can tolerate a wide variety of side chains, including hydrophilic and hydrophobic residues. This result might be taken to indicate that surface positions contain little structural information. However, Bashford *et al.*, in an extensive analysis of globin sequences (*4*), found a strong bias against large hydrophobic residues at many surface positions. At one level, this may reflect constraints imposed by protein solubility, because large patches of hydrophobic surface residues would presumably lead to aggregation. At a more fundamental level, protein folding requires a partitioning between surface and buried positions. Consequently, to achieve a unique native state without significant competition from other conformations, it may be important that some sites have a decided preference for exterior rather than interior positions. As a result, many surface sites can accept hydrophobic residues individually, but the surface as a whole can probably tolerate only a moderate number of hydrophobic side chains.

Identification of Residue Roles from Sets of Sequences

Often, a protein of interest is a member of a family of related sequences. What can we infer from the pattern of allowed substitutions at positions in sets of aligned sequences generated by genetic or phylogenetic methods? Residue positions that can accept a number of different side chains, including charged and highly polar residues, are almost certain to be on the protein surface. Residue positions that remain hydrophobic, whether variable or not, are likely to be buried within the structure. In Fig. 3, those residue positions in λ repressor that can accept hydrophilic side chains are shown in orange and those that cannot accept hydrophilic side chains are shown in green. The obligate hydrophobic positions define the core of the structure, whereas positions that can accept hydrophilic side chains define the surface.

Functionally important residues should be conserved in sets of active sequences, but it is not possible to decide whether a side chain s functionally or structurally important just because it is invariant or conserved. To make this distinction requires an independent assay of protein folding. The ability of a mutant protein to maintain a stably folded structure can often be measured by biophysical techniques, by susceptibility to intracellular proteolysis (*26*), or by binding to antibodies specific for the native structure (*27, 28*). In the latter cases, it is possible to screen proteins in mutated clones for the ability to fold even if these proteins are inactive. Sets of sequences that allow formation of a stable structure can then be compared to the sets that allow both folding and function, with the active site or binding residues being those that are variable in the set of stable proteins but invariant in the set of functional proteins. The DNA-binding residues of Arc repressor were identified by this method (*8*). The receptor-binding residues of human growth hormone were also identified by comparing the stabilities and activities of a set of mutant sequences (*28*). However, in this case, the mutants were generated as hybrid sequences between growth hormone and related hormones with different binding specificities.

Implications for Structure Prediction

At present, the only reliable method for predicting a low-resolution tertiary structure of a new protein is by identifying sequence similarity to a protein whose structure is already known (*29, 30*). However, it is often difficult to align sequences as the level of sequence similarity decreases, and it is sometimes impossible to detect statistically significant sequence similarity between distantly related proteins. Because the number of known sequences is far greater than the number of known structures, it would be advantageous to increase the reach of the available structural information by improving methods for detecting distant sequence relations and for subsequently aligning these sequences based on structural principles. In a normal homology search, the sequence database is scanned with a single test sequence, and every residue must be weighted equally. However, some residues are more important than others and should be weighted accordingly. Moreover, certain regions of the protein are more likely to contain gaps than others. Both kinds of information can be obtained from sequence sets, and several techniques have

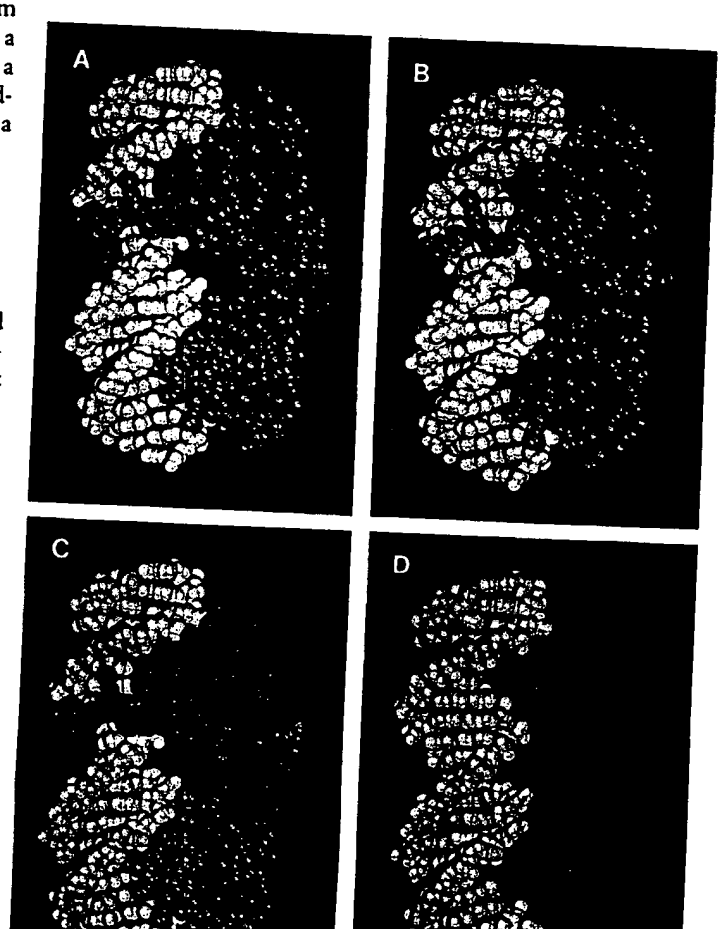

Fig. 3. Tolerance of positions in the NH_2-terminal domain of λ repressor to hydrophilic side chains. The complex (*43*) of the repressor dimer (blue) and operator DNA (white) is shown. In (**A**), positions that can tolerate hydrophilic side chains are shown in orange. The same side chains are shown in (**B**) without the remaining protein atoms. In (**C**), positions that require hydrophobic or neutral side chains are shown in green. These side chains are shown in (**D**) without the remaining protein atoms. About three-fourths of the 92 side chains in the NH_2-terminal domain are included in both (B) and (D). The remaining positions have not been tested. Data are from (*9, 14, 20, 27, 44*).

been used to combine such information into more appropriately weighted sequence searches and alignments (*31*). These methods were used to align the sequences of retroviral proteases with aspartic proteases, which in turn allowed construction of a three-dimensional model for the protease of human immunodeficiency virus type 1 (*29*). Comparison with the recently determined crystal structure of this protein revealed reasonable agreement in many areas of the predicted structure (*32*).

The structural information at most surface sites is highly degenerate. Except for functionally important residues, exterior positions seem to be important chiefly in maintaining a reasonably polar surface. The information contained in buried residues is also degenerate, the main requirement being that these residues remain hydrophobic. Thus, at its most basic level, the key structural message in an amino acid sequence may reside in its specific pattern of hydrophobic and hydrophilic residues. This is meant in an informational sense. Clearly, the precise structure and stability of a protein depends on a large number of detailed interactions. It is possible, however, that structural prediction at a more primitive level can be accomplished by concentrating on the most basic informational aspects of an amino acid sequence. For example, amphipathic patterns can be extracted from aligned sets of sequences and used, in some cases, to identify secondary structures.

If a region of secondary structure is packed against the hydrophobic core, a pattern of hydrophobic residues reflecting the periodicity of the secondary structure is expected (*33, 34*). These patterns can be obscured in individual sequences by hydrophobic residues on the protein surface. It is rare, however, for a surface position to remain hydrophobic over the course of evolution. Consequently, the amphipathic patterns expected for simple secondary structures can be much clearer in a set of related sequences (*6*). This principle is illustrated in Fig. 4, which shows helical hydrophobic moment plots for the Antennapedia homeodomain sequence (Fig. 4A) and for a composite sequence derived from a set of homologous homeodomain proteins (Fig. 4B) (*35*). The hydrophobic moment is a simple measure of the degree of amphipathic character of a sequence in a given secondary structure (*34*). The amphipathic character of the three α-helical regions in the Antennapedia protein (*36*) is clearly revealed only by the analysis of the combined set of homeodomain sequences. The secondary structure of Arc repressor, a small DNA-binding protein, was recently predicted by a similar method (*8*) and confirmed by nuclear magnetic resonance studies (*37*).

The specific pattern of hydrophobic and hydrophilic residues in an amino acid sequence must limit the number of different structures a given sequence can adopt and may indeed define its overall fold. If this is true, then the arrangement of hydrophobic and hydrophilic residues should be a characteristic feature of a particular fold. Sweet and Eisenberg have shown that the correlation of the pattern of hydrophobicity between two protein sequences is a good criterion for their structural relatedness (*38*). In addition, several studies indicate that patterns of obligatory hydrophobic positions identified from aligned sequences are distinctive features of sequences that adopt the same structure (*4, 29, 38, 39*). Thus, the order of hydrophobic and hydrophilic residues in a sequence may actually be sufficient information to determine the basic folding pattern of a protein sequence.

Although the pattern of sequence hydrophobicity may be a characteristic feature of a particular fold, it is not yet clear how such patterns could be used for prediction of structure de novo. It is important to understand how patterns in sequence space can be related to structures in conformation space. Lau and Dill have approached this problem by studying the properties of simple sequences composed only of H (hydrophobic) and P (polar) groups on two-dimensional lattices (*40*). An example of such a representation is shown in Fig. 5. Residues adjacent in the sequence must occupy adjacent squares on the lattice, and two residues cannot occupy the same space. Free energies of particular conformations are evaluated with a single term, an attraction of H groups. By considering chains of ten residues, an exhaustive conformational search for all 1024 possible sequences of H and P residues was possible. For longer sequences only a representative fraction of the allowed sequence or conformation space could be explored. The significant results were as follows: (i) not all sequences can fold into a "native" structure and only a few sequences form a unique native structure; (ii) the probability that a sequence will adopt a unique native structure increases with chain length; and (iii) the native states are compact, contain a hydrophobic core surrounded by polar residues, and contain significant secondary structure. Although the gap between these two-dimensional simulations and three-dimensional structures is large, the use of simple rules and sequence representations yields results similar to those expected for real proteins. Three-dimensional lattice methods are also beginning to be developed and evaluated (*41*).

Summary

There is more information in a set of related sequences than in a single sequence. A number of practical applications arise from an analysis of the tolerance of residue positions to change. First, such information permits the evaluation of a residue's importance to the function and stability of a protein. This ability to identify the essential elements of a protein sequence may improve our understanding of the determinants of protein folding and stability as well as protein function. Second, patterns of tolerance to amino acid substitutions of varying hydrophilicity can help to identify residues likely to be buried in a protein structure and those likely to occupy

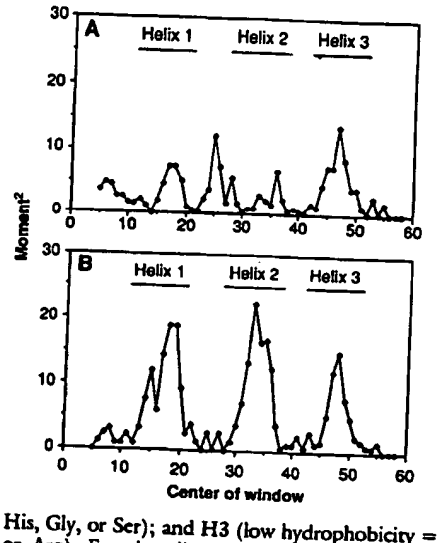

Fig. 4. Helical hydrophobic moments calculated by using (**A**) the Antennapedia homeodomain sequence or (**B**) a set of 39 aligned homeodomain sequences (*35*). The bars indicate the extent of the helical regions identified in nuclear magnetic resonance studies of the Antennapedia homeodomain (*36*). To determine hydrophobic moments, residues were assigned to one of three groups: H1 (high hydrophobicity = Trp, Ile, Phe, Leu, Met, Val, or Cys); H2 (medium hydrophobicity = Tyr, Pro, Ala, Thr, His, Gly, or Ser); and H3 (low hydrophobicity = Gln, Asn, Glu, Asp, Lys, or Arg). For the aligned homeodomain sequences, the residues at each position were sorted by their hydrophobicity by using the scale of Fauchere and Pliska (*45*). Arg and Lys were not counted unless no other residue was found at the position, because they contain long aliphatic side chains and can thereby substitute for nonpolar residues at some buried sites. To account for possible sequence errors and rare exceptions, the most hydrophilic residue allowed at each position was discarded unless it was observed twice. The second most hydrophilic residue was then chosen to represent the hydrophobicity of each position. An eight-residue window was used and the vectors projected radially every 100°. The vector magnitudes were assigned a value of 1, 0, or −1 for positions where the hydrophobicity group was H1, H2, or H3, respectively.

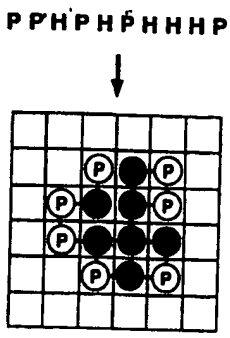

Fig. 5. A representation of one compact conformation for a particular sequence of H and P residues on a two-dimensional square lattice. [Adapted from (*40*), with permission of the American Chemical Society]

surface positions. The amphipathic patterns that emerge can be used to identify probable regions of secondary structure. Third, incorporating a knowledge of allowed substitutions can improve the ability to detect and align distantly related proteins because the essential residues can be given prominence in the alignment scoring.

As more sequences are determined, it becomes increasingly likely that a protein of interest is a member of a family of related sequences. If this is not the case, it is now possible to use genetic methods to generate lists of allowed amino acid substitutions. Consequently, at least in the short term, it may not be necessary to solve the folding problem for individual protein sequences. Instead, information from sequence sets could be used. Perhaps by simplifying sequence space through the identification of key residues, and by simplifying conformation space as in the lattice methods, it will be possible to develop algorithms to generate a limited number of trial structures. These trial structures could then, in turn, be evaluated by further experiments and more sophisticated energy calculations.

REFERENCES AND NOTES

1. C. J. Epstein, R. F. Goldberger, C. B. Anfinsen, *Cold Spring Harbor Symp. Quant. Biol.* **28**, 439 (1963); C. B. Anfinsen, *Science* **181**, 223 (1973).
2. R. E. Dickerson, *Sci. Am.* **242**, 136 (March 1980).
3. M. D. Hampsey, G. Das, F. Sherman, *FEBS Lett.* **231**, 275 (1988).
4. D. Bashford, C. Chothia, A. M. Lesk, *J. Mol. Biol.* **196**, 199 (1987).
5. A. M. Lesk and C. Chothia, *ibid.* **136**, 225 (1980).
6. M. F. Perutz, J. C. Kendrew, H. C. Watson, *ibid.* **13**, 669 (1965).
7. C. Chothia and A. M. Lesk, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 399 (1965).
8. J. U. Bowie and R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2152 (1989).
9. J. F. Reidhaar-Olson and R. T. Sauer, *Science* **241**, 53 (1988); *Proteins Struct. Funct. Genet.*, in press.
10. D. Shortle, *J. Biol. Chem.* **264**, 5315 (1989).
11. J. H. Miller *et al.*, *J. Mol. Biol.* **131**, 191 (1979).
12. S. Sprang *et al.*, *Science* **237**, 905 (1987); C. S. Craik, S. Rozcniak, C. Largman, W. J. Rutter, *ibid.*, p. 909.
13. H. C. M. Nelson and R. T. Sauer, *J. Mol. Biol.* **192**, 27 (1986).
14. M. H. Hecht, J. M. Sturtevant, R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5685 (1984).
15. T. Alber, D. Sun, J. A. Nye, D. C. Muchmore, B. W. Matthews, *Biochemistry* **26**, 3754 (1987).
16. D. Shortle and A. K. Meeker, *Proteins Struct. Funct. Genet.* **1**, 81 (1986).
17. A. M. Lesk and C. Chothia, *J. Mol. Biol.* **160**, 325 (1982).
18. W. R. Taylor, *ibid.* **188**, 233 (1986).
19. W. Kauzmann, *Adv. Protein Chem.* **14**, 1 (1959); R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8069 (1986).
20. W. A. Lim and R. T. Sauer, *Nature* **339**, 31 (1989); in preparation.
21. Lesk and Chothia (*5*) have argued that a protein core composed solely of hydrogen-bonded residues would also be inviable on evolutionary grounds, as a mutational change in one core residue would require compensating changes in any interacting residue or residues to maintain a stable structure.
22. T. M. Gray and B. W. Matthews, *J. Mol. Biol.* **175**, 75 (1984); E. N. Baker and R. E. Hubbard, *Prog. Biophys. Mol. Biol.* **44**, 97 (1984).
23. F. M. Richards, *J. Mol. Biol.* **82**, 1 (1974).
24. J. W. Ponder and F. M. Richards, *ibid.* **193**, 775 (1987).
25. J. T. Kellis, Jr., K. Nyberg, A. R. Fersht, *Biochemistry* **28**, 4914 (1989); W. S. Sandberg and T. C. Terwilliger, *Science* **245**, 54 (1989).
26. A. A. Pakula and R. T. Sauer, *Proteins Struct. Funct. Genet.* **5**, 202 (1989).
27. B. C. Cunningham and J. A. Wells, *Science* **244**, 1081 (1989); R. M. Breyer and R. T. Sauer, *J. Biol. Chem.* **264**, 13348 (1989).
28. B. C. Cunningham, P. Jhurani, P. Ng, J. A. Wells, *Science* **243**, 1330 (1989).
29. L. H. Pearl and W. R. Taylor, *Nature* **329**, 351 (1987).
30. W. J. Brown *et al.*, *J. Mol. Biol.* **42**, 65 (1969); J. Greer, *ibid.* **153**, 1027 (1981); J. M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 99 (1988).
31. W. R. Taylor, *Protein Eng.* **2**, 77 (1988).
32. M. A. Navia *et al.*, *Nature* **337**, 615 (1989).
33. M. Schiffer and A. B. Edmundson, *Biophys. J.* **7**, 121 (1967); V. I. Lim, *J. Mol. Biol.* **88**, 857 (1974); *ibid.*, p. 873.
34. D. Eisenberg, R. M. Weiss, T. C. Terwilliger, *Nature* **299**, 371 (1982); D. Eisenberg, D. Schwarz, M. Komaromy, R. Wall, *J. Mol. Biol.* **179**, 125 (1984); D. Eisenberg, R. M. Weiss, T. C. Terwilliger, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 140 (1984).
35. T. R. Burglin, *Cell* **53**, 339 (1988).
36. G. Otting *et al.*, *EMBO J.* **7**, 4305 (1988).
37. J. N. Breg, R. Boelens, A. V. E. George, R. Kaptein, *Biochemistry* **28**, 9826 (1989); M. G. Zagorski, J. U. Bowie, A. K. Vershon, R. T. Sauer, D. J. Patel, *ibid.*, p. 9813.
38. R. M. Sweet and D. Eisenberg, *J. Mol. Biol.* **171**, 479 (1983).
39. J. U. Bowie, N. D. Clarke, C. O. Pabo, R. T. Sauer, *Proteins Struct. Funct. Genet.*, in preparation.
40. K. F. Lau and K. A. Dill, *Macromolecules* **22**, 3986 (1989).
41. A. Sikorski and J. Skolnick, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2668 (1989); A. Kolinski, J. Skolnick, R. Yaris, *Biopolymers* **26**, 937 (1987); D. G. Covell and R. L. Jernigan, *Biochemistry*, in press.
42. B. Lee and F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971).
43. S. R. Jordan and C. O. Pabo, *Science* **242**, 893 (1988).
44. R. M. Breyer, thesis, Massachusetts Institute of Technology, Cambridge (1988).
45. J.-L. Fauchere and V. Pliska, *Eur. J. Med. Chem.-Chim. Ther.* **18**, 369 (1983).

46. We thank C. O. Pabo and S. Jordan for coordinates of the NH_2-terminal domain of λ repressor and its operator complex. We also thank P. Schimmel for the use of his graphics system and J. Burnbaum and C. Francklyn for assistance. Supported in part by NIH grant AI-15706 and predoctoral grants from NSF (J.R.-O.) and Howard Hughes Medical Institute (W.A.L.).